UCL

DIVISION OF MEDICINE

SCHOOL OF LIFE AND MEDICAL SCIENCES

Structure-based drug design for the discovery of new treatments for trypanosomiasis

A submission for the degree of Doctor of Philosophy

Eyram Adjogatse

September 2014

I. Declaration

I, Eyram Adjogatse confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed,

_____

Eyram Adjogatse

II. Abstract

Human African trypanosomiasis (HAT) and Chagas disease are caused by infection with the protozoan parasites *Trypanosoma brucei* and *T. cruzi*, respectively. There has historically been a lack of investment into measures to control these diseases. As a result, few drugs are available to treat HAT and Chagas disease, and there is an urgent need for novel alternatives.

The enzyme L-threonine 3-dehydrogenase (TDH) initiates the conversion of L-threonine into acetyl-coenzyme A and glycine. This pathway has been shown to play a vital role in *T. brucei*, particularly in fatty acid synthesis. Exposure of *T. brucei* in culture to a potent TDH inhibitor, has been shown to be lethal(1) and dual blockade of the TDH pathway and a second pathway for acetyl-coenzyme A production, terminated by pyruvate dehydrogenase, completely inhibits the growth of *T. brucei*(2,3).

Multiple three-dimensional structures of TDH, alone and in complex with ligands, were determined by X-ray crystallography. In parallel, enzyme assays were carried out to investigate the kinetic behaviour of TDH and the modes of action of known TDH inhibitors. The structural information on TDH was used in a virtual screen to predict the binding interactions between the enzyme and a library of around 3000 ligands. These ligands were mainly selected for their diversity and for their inhibition of proteins related to TDH. Subsequently, an *in vitro* screen was performed to test compounds identified by virtual screening, along with small molecules and fragments from commercial libraries.

In total, 27 compounds were identified as TDH inhibitors. Of these compounds, four were found to potently inhibit *T. brucei* growth. This study has demonstrated the effectiveness of combining structural and functional data in rational drug discovery. Novel aspects of TDH have been discovered, in addition to novel inhibitors that will aid in the design of a new class of antitrypanosomal drugs.

References:

1. Cross GA, Klein RA, Linstead DJ. Utilization of amino acids by *Trypanosoma brucei* in culture: L-threonine as a precursor for acetate. Parasitology. 1975 Oct;71(2):311–26.
2. Millerioux Y, Ebikeme C, Biran M, Morand P, Bouyssou G, Vincent IM, et al. The threonine degradation pathway of the *Trypanosoma brucei* procyclic form: the main carbon source for lipid biosynthesis is under metabolic control. Mol Microbiol. 2013 Oct;90(1):114–29.
3. Mazet M, Morand P, Biran M, Bouyssou G, Courtois P, Daulouède S, et al. Revisiting the Central Metabolism of the Bloodstream Forms of *Trypanosoma brucei*: Production of Acetate in the Mitochondrion Is Essential for Parasite Viability. PLoS Negl Trop Dis. 2013 Dec 19;7(12):e2587.

IV. Table of contents

V. List of figures

VI. List of tables

VII. List of abbreviations

ADT	AutoDock Tools	**DLiS**	Dynamic Light Scattering
AIC	Aikake's Information Criterion	**DLS**	Diamond Light Source
AKB	2-amino-3-ketobutyrate	**DML**	Drug multiple ligands
AMC	Advance Purchase/Market Commitment	**DMS**	Dimethyl suberimidate
ASAQ	Artesunate and amodiaquine	**DMSO**	Dimethyl sulfoxide
ASMQ	Artesunate and mefloquine	**DNA**	Deoxyribonucleic acid
AUC	Analytical ultracentrifugation	**DNDi**	Drugs for Neglected Diseases initiative
BLAST	Basic Local Alignment Search Tool	**DTP**	NCI/NIH Developmental Therapeutics Program
βM	β-mercaptoethanol	**DVMS**	Data Visualisation Modelling System
BPOB	2-{[3-(4-bromophenyl)-3-oxopropyl]amino}-3-hydroxybutanoic acid	**ESRF**	European Synchrotron Radiation Facility
BSA	Bovine serum albumin	**FDC**	Fixed Dose Combination
BSF	Bloodstream form (trypanosomes)	***Ff*TDH**	TDH from *Flavobacterium frigidimaris*
BVGH	BIO Ventures for Global Health	**FPLC**	Fast Performance Liquid Chromatography
***Cd*TDH**	TDH from *Clostridium difficile*	**GalE**	Uridine diphosphate-galactose 4'-epimerase
clogP	Calculated log P	**GDP**	Guanosine diphosphate
***Cn*TDH**	TDH from *Cupriavidus necator*	**G-FINDER**	Global Funding for Innovation in Neglected Diseases
CoA	Coenzyme A	**GlmS**	Glucosamine-6-phosphate synthase
CRO	Contract Research Organisation	**GUI**	Graphical user interface
DFID	Department for International Development	**HAT**	Human African trypanosomiasis
DFR	Dihydroflavonol (Dihydrokaempferol) 4'-reductase	**IC_{50}**	Concentration of inhibitor producing 50% inhibition

IC_{90}	Concentration of inhibitor producing 90% inhibition	**NAD^+**	Nicotinamide adenine dinucleotide - oxidised form
IDRI	Infectious Disease Research Institute	**NADH**	Nicotinamide adenine dinucleotide - reduced form
IHME	Institute for Health Metrics Evaluation	**NCBI**	National Center for Biotechnology Information
IOWH	Institute for One World Health	**NCE**	New Chemical Entity
IPTG	isopropyl ß-D-thiogalactopyranoside	**NCI**	National Cancer Institute
KBL	2-Amino-3-ketobutyrate ligase	**NF**	New Formulation
k_{cat}	Turnover number	**NGO**	Non-governmental Organisation
K_i	Inhibitory constant	**NI**	New Indication
k_{inact}	Rate of enzyme inactivation (covalent/irreversible)	**NIH**	US National Institutes of Health
K_M	Half-maximal/half-saturating substrate concentration; Michaelis constant	**NME**	New Molecular Entity
KNF	Koshland-Nemethy-Filmer/Simple sequential interaction model of cooperativity	**NTD**	Neglected Tropical Disease
k_{obs}	Rate of enzyme inhibition (time-dependent/slow-binding)	**OD**	Optical density
LB	Luria Broth	**PDB**	Protein Data Bank
logP	Log octanol:water partition coefficient (P)	**PRV**	Priority Review Voucher
LSHTM	London School of Hygiene and Tropical Medicine	**PSA**	Polar surface area
MA	Marketing Authorisation	**PSNRI**	Public-Sector or Not-for-profit Research Institute
***Mm*TDHi**	Inhibitor of TDH from *Mus musculus* (mouse)	**PSSC**	Protein Structure Similarity Clustering
MW	Molecular weight	**PUT**	Plate uniformity test
MWC	Monod-Wyman-Changeux	**Qc**	Quinazoline carboxamide
n.d.	Not determined	**R&D**	Research and Development
NAD (NAD(H))	Nicotinamide adenine dinucleotide	**RMSD**	Root-mean-square deviation

RNA	ribonucleic acid
RNAi	RNA interference
S&T	Science and Technology
SAR	Structure-activity relationship
SD	Standard deviation
SD	Standard deviation
SDR	Short-chain dehydrogenase/reductase
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SIB	Social Impact Bond
SlyD	Peptidoylproline cis–trans isomerase
SMILES	Simplified Molecular Input Line Entry System (representation of chemical structure)
SoP	UCL School of Pharmacy
TbGalE	GalE from Trypanosoma brucei
***Tb*TDH**	TDH from *Trypanosoma brucei*
TDH	L-threonine 3-dehydrogenase
TDP	Thymidine diphosphate
TETD	Tetra ethyl thiuram disulfide
TPP	Target Product Profile
TPSA	Topological polar surface area
***Tb*GalE**	GalE from Trypanosoma brucei
***Tb*TDH**	TDH from Trypanosoma brucei
TDH	L-threonine 3-dehydrogenase
TDP	Thymidine diphosphate
TETD	Tetra ethyl thiuram disulfide
TPP	Target Product Profile
TPSA	Topological polar surface area
***Tv*TDH**	TDH from *Thermoplasma volcanium*
UDP	uridine diphosphate
UMP	uridine monophosphate
USAID	United States Agency for International Development
V_0	Void volume
V_e	Elution volume
VL	Visceral Leishmaniasis
V_{max}	Maximum velocity of enzyme activity
VS	Virtual Screen
WHO	World Health Organisation
WHO EML	WHO Essential Medicines List
WHO PQ	WHO Pre-qualified medicine
WHO TDR	WHO Special Programme for Research and Training in Tropical Diseases

1. Introduction

1.1 Rational Drug Discovery

Over the past century an unprecedented increase in the longevity and quality of human life has occurred. That increase has been claimed by many to be largely influenced by the discovery, development and distribution of pharmaceuticals(1–3). Steady increases in the development of new drugs in the second half of the twentieth century were accompanied by an increasing use of rational methods of drug discovery over empirical ones(4–9). Rational drug discovery is the discovery and design of drugs based on information of a particular target molecule, typically a protein, that has been implicated in disease(4,5). Certainly, the efficacy of this approach has been augmented by the introduction of several new technologies over the years. Firstly, the completion of the human genome project and the completion of genomes for other organisms has created access to a huge repertoire of "druggable" targets(10–13). Structural Biology, coupled with computational methods has been one of the greatest enablers of rational drug design. In recent years, the increase in high-throughput technologies in chemistry and biology have driven the success of the field(5).

The debate over whether rational target-based approaches or empirical approaches, which focus principally on the clinical and phenotypical effects of investigational drugs, are more effective for drug discovery is ongoing. Drug candidates discovered through rational approaches are frequently unable to reproduce effects demonstrated *in vitro* once used in the clinical setting. Furthermore, if a particular target has not been well validated (see Figure 1.1.1, 'Target Discovery & Validation') then drugs affecting that target will fail to cause the desired effect. This has been cited as a cause of several late-stage drug failures(14). On the other hand, in empirical approaches to drug discovery, drugs are discovered and optimised based on their ability to produce a desired effect. Although this seems like more of a 'blind' approach to drug discovery, it can avoid many of the problems associated with rational drug discovery. However, a lack of knowledge of a drug's mechanism of action can mean that important events, such as side-effects and drug resistance, are more difficult or sometimes impossible to predict. Although empirical or phenotypic drug discovery is experiencing a renaissance, culturally, there is still an expectation that the mechanism of action of a drug be known before it is approved for use. In the case of certain medical conditions, certain cell types will need to be targeted. Therefore, rational drug discovery and design is often favoured, particularly in anti-cancer drug discovery and in anti-infective drug discovery, where differences between the pathogen and host physiology can be exploited.

Rational, structure-based drug design often proceeds along two arms – target-based investigations and ligand-based investigations – with an optimal interaction between the target and a ligand being the end-goal. Target-based investigations focus on the characterisation of

the target's function and on other characteristics, such as the target's structure. Ligand-based investigations tend to focus on the characteristics of ligands and on how these characteristics are associated with the ligand's ability to modulate a target's function. Gaining information on the structure of a target and on how its structure relates to its function can be crucial to the advancement of a drug discovery project. Structural studies feature a number of methods, including electron microscopy, nuclear magnetic resonance, and most prominently, X-ray crystallography. X-ray crystallography can produce models of the molecular structure of a crystallised substance from the diffraction pattern produced by it upon exposure to X-ray radiation. X-ray crystallography has had an extremely important role in drug design as it has been the way in which the most detailed and definitive data on small molecule or macromolecular structures have been obtained, and it has guided many successful lead development efforts(5,15–18).

Ligand-based methods, those focusing on the chemical structure of active ligands and the development of structure activity relationships (SARs), are often used in the absence of structural information on the target, but find their most powerful application when combined with target information(9). When an active ligand is discovered, it can be highly valuable to elucidate the molecular interactions between it and the target. In that way, ligand effects on the function of the target can be linked to interactions with specific structural features of the target. Then, the SARs of different ligands can be expanded to include activities related to specific molecular interactions(9,19,20). X-ray crystallography is particularly useful for investigating ligand-macromolecule interactions, due to its ability to determine structures at atomic or near-atomic resolution.

Screening is used frequently in target-based approaches. Although often criticised as not being effective in isolation, it can form a vital part of a rational drug design program(5). The range of assays now used in screening experiments includes cell-based, biochemical and even biophysical assays. Screening typically involves the testing of a large and diverse library of chemicals in the hope that a number of active compounds (hits) will be discovered. Alternatively, a 'focused' library of compounds related to a known active ligand may be screened. The scale of screening can vary significantly, from a few thousand compounds to as many as millions of compounds, which would usually be achieved by high-throughput screening (HTS). The hits identified in a screening experiment can serve many purposes. Firstly, the hits may be used for further study of the target and for the development of SARs linked to molecular interactions. Knowledge of specific pharmacophores, the features of a hit that are responsible for its effect, can then be used in the *de novo* design of new ligands. In other situations, depending on the hits which are identified, an active compound can serve as a starting point for the design of a new drug. This compound, a lead, is then chemically optimised to improve its activity and to alter its properties to make it more suitable as a drug. In particular, properties that may affect the compound's pharmacokinetics (absorption, distribution, metabolism,

excretion and toxicity) in man are considered imperative. Fragment-based drug discovery makes use of SARs established by studying several ligands and also uses hits as starting points for lead design. The fragment-based approach screens smaller active compounds (molecular weight ~100-250 Da) than those traditionally used in screening (molecular weight ~300-600 Da). Accordingly, fragments are usually of lower potency than higher-molecular weight hits. However, fragments binding to different, adjacent parts of the target molecule can be linked together to form more active compounds. Alternatively, the same outcome can be achieved by "growing" a fragment, by adding new functional groups to the original hit compound. Building a lead from a low-molecular weight compound has allowed medicinal chemists to design potent leads, whilst maintaining good physicochemical properties of the compound(21,22). Fragment-based drug discovery is now at the centre of many academic and industrial drug discovery programs, and it is often used in combination with more traditional approaches. The impact of this approach on drug discovery is becoming evident, as several drugs discovered by fragment-based discovery are now in the clinic or late-stage development(21).

Computational chemistry has grown as a discipline due to its ability to integrate data from several sources, such as X-ray crystallography and HTS, and to allow the testing and development of theories virtually before proceeding with *in vitro* studies(9,23). Molecular docking programs, which attempt to predict interactions between a ligand and a target are particularly popular tools in computational drug discovery(24–29). A great advantage of including such methods in drug discovery projects is that they can speed up the process and improve efficiency(9,30), thus expanding the range of environments in which such activities can take place(31). This is particularly useful where high-throughput facilities are unavailable and screening large screening libraries (e.g. hundreds of thousands of compounds) is not feasible. Potential hit compounds can be identified *in silico* by a virtual screening experiment using molecular docking software. Then, a focused library of virtual screening hits can be assayed in an *in vitro* screen, therefore reducing the number of molecules that need to be screened to identify hits.

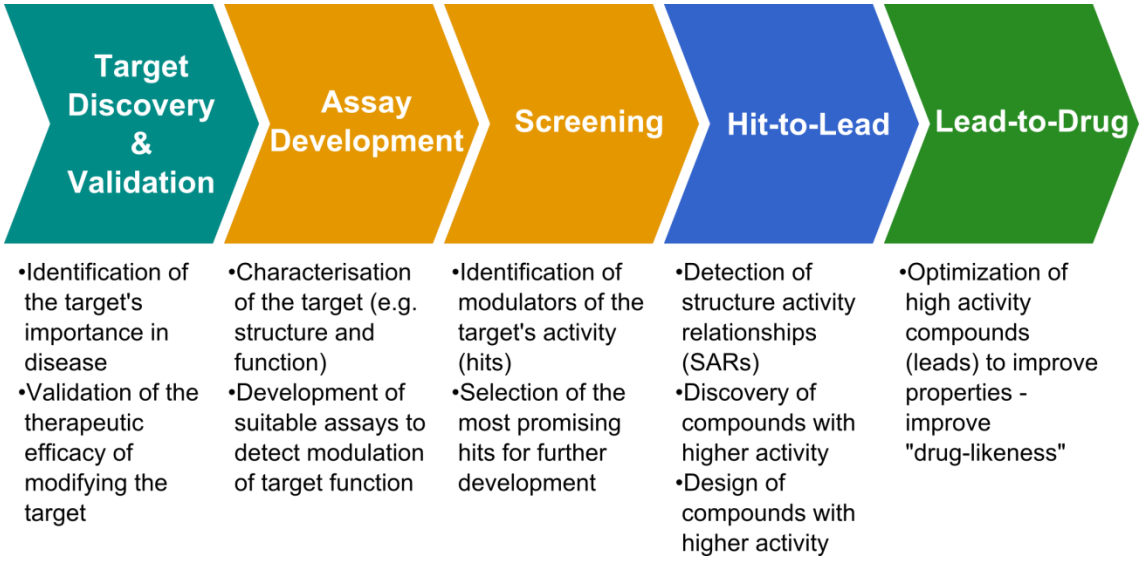


Figure 1.1.1 - early-stage drug discovery and the steps to the development of a new investigational drug. The steps represented by gold coloured arrows are those involved in the work included in this thesis.

In recent years, the concept of "druggability", which describes the ability of a target to be modulated by a drug, has arisen. In one respect, a target's druggability is affected by the ability of a drug, when administered, to accumulate to an efficacious concentration in the environment of the target. Therefore, the tissue and cellular location of a target is very important. In another respect, particularly important at the earliest stages of drug discovery, a target is considered more druggable if drug-like molecules are able to bind to it with high affinity and affect its activity. Therefore, the structural and physicochemical features of a protein are essential in determining its suitability as a drug target(32–35). This is exemplified by the fact that G-protein-coupled receptors and enzymes are the most common targets of marketed drugs(10). Many of these proteins have evolved to bind small-molecules, and so possess protein folds and binding sites suitable for high-affinity ligand binding. Other factors influencing the selection of a drug target are related to the therapeutic area in which a drug is being developed.

1.2 Trypanosomiasis and Neglected Tropical Diseases (NTDs)

Neglected Tropical Diseases (NTDs) are mainly prevalent in developing nations with tropical climates. They primarily affect the poor in these countries and therefore have received little investment from the pharmaceutical industry into new treatments(36,37). Following the previous "neglect" of these diseases, both in research strategy and global health policy(38), there has been renewed interest in NTDs, following the coining of the phrase in 2005(36), and an increased prominence of NTDs in global/public health strategies(39–43). The World Health Organisation (WHO) lists 17 NTDS, including: Onchocerciacis, Leishmaniasis, Schistosomiasis, Chagas disease, Human African Trypanosomiasis, Dengue Fever and Yaws, amongst

others(44). NTDs contribute a particularly large part of the disease burden in the developing world and they perpetuate the poverty and underdevelopment of these regions(41,45,46).

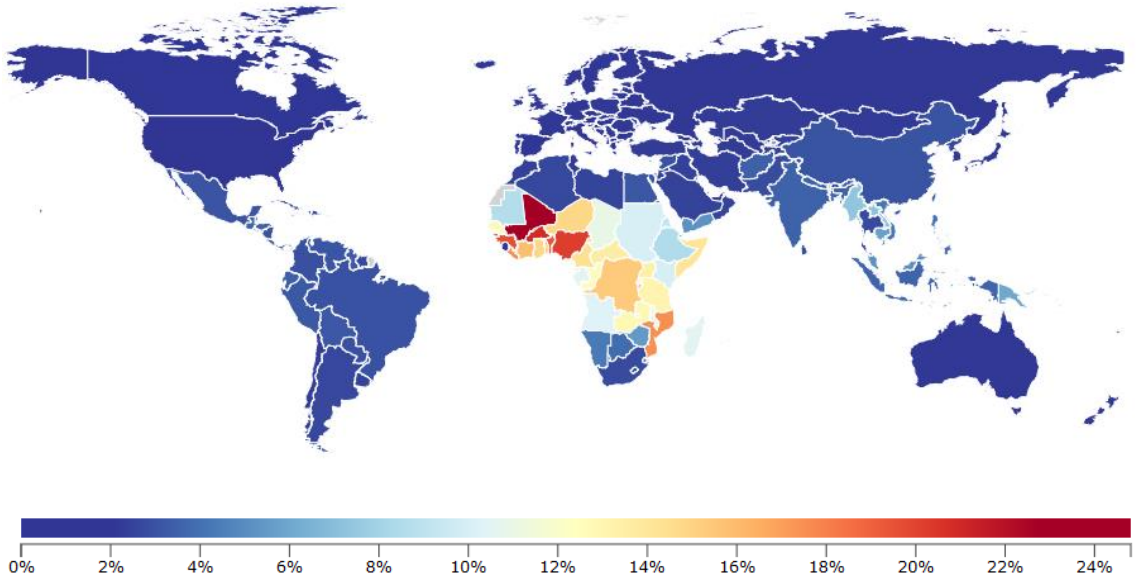


Figure 1.2.1 - the global distribution of the burden of NTDs in 2010. Countries are coloured according to the percentage of the total disability-adjusted life years (DALYs) caused NTDs in that country. NTDs are most prevalent in sub-Saharan Africa, Latin America & the Caribbean, and in South & Southeast Asia. The map was attained using the 'GBD Compare' tool of the Institute of Health Metrics and Evaluation (IHME), University of Washington(47).

1.2.1 The NTD health burden

NTDs collectively represent an enormous health burden globally. These diseases are infectious diseases, generally concentrated in the Global South. The WHO has defined a list of 17 NTDs(48), whilst a number of other diseases have been identified as NTDs by other organisations(41,49) (see Figure 1.2.2). It has been estimated that NTDs represent 10-12% of the global disease burden(50–52), but receive a relatively small investment in R&D. NTDs are now often referred to as diseases of "neglected populations", as they often exist in communities that are socially excluded, sometimes due to their geographic location, but also as a result of their impoverished condition, to which infectious diseases contribute(41). As a result of this situation, NTDs lie as a barrier to economic development and represent a significant expenditure to governments in disease-endemic areas (41). For example, an estimated US$267million is spent annually on the treatment of Chagas Disease patients in Colombia alone(53). Thus with such economic and social costs resulting from NTDs globally, it is within the interests of private, public and philanthropic organisations, both nationally and internationally, to bring them under control.

"The Big Three"
- HIV/AIDS
- Malaria
- Tuberculosis

Protozoan Infections

- Amebiasis
- Balantidiasis
- Chagas Disease*
- Giardiasis
- Human African Trypanosomiasis*
- Leishmaniasis*

Helminth Infections

- Taeniasis-Cysticercosis*
- Dracunculiasis*
- Echinococcosis
- Food-borne Trematodiases*
- Loiasis
- Lymphatic Filariasis*
- Onchocerciasis*
- Schistosomiasis*
- Soil-transmitted Helminthiases (Ascariasis, Hookworm Diseases, Trichuriasis, Strongyloidiasis)*
- Toxocariasis and other Larva Migrans

Other

- Snakebite

Viral Infections

- Dengue*
- Japanese encephalitis
- Jungle yellow fever
- Other arboviral infections
- Rabies*
- Rift Valley fever
- Viral hemorrhagic fevers

Bacterial Infections

- Bartonella
- Bovine Tuberculosis in Humans
- Buruli Ulcer*
- Cholera
- Enteric pathogens (Shigella, Salmonella, E. coli)
- Leprosy*
- Leptospirosis
- Relapsing Fever
- Trachoma*
- Treponematoses (Bejel, Pinta, Syphilis, Yaws*)

Fungal Infections

- Mycetoma
- Paracoccidiomycosis

Ectoparasitic Infections

- Scabies
- Myiasis

Figure 1.2.2 - Diseases classified as NTDs. This list was constructed based on definitions by Hotez et al.(41) and the scope of the journal PLOS Neglected Tropical Diseases(49). "The Big Three" are excluded from some definitions in the literature, as they receive much more attention than other NTDs. * indicates a disease included in the WHO list of 17 core neglected tropical diseases.

1.2.2 NTD policy

Historical

In the disease endemic countries, many NTDs have been present for several thousands of years, periodically presenting as epidemics with several fatalities. The human and economic costs of NTDs were recognised by the colonial powers of former colonies, which led to the implementation of mass screening, treatment and vector control programs(50,54). In fact, many of the most burdensome NTDs once posed barriers to economic progress in Western countries, and their subsequent control and elimination increased the prosperity of affected areas. However after several of the former colonies gained independence and disease prevalence had decreased, control programs were relaxed and some of these diseases began to re-emerge.

Since that time, various famines, regional conflicts and political crises have exacerbated the situation.

Awareness and policy

There are a number of reasons for the persistence of today's NTDs. The WHO have described how certain diseases lack the appropriate "tools" (e.g. drugs, vaccines, diagnostics) to control them, whereas with other NTDs, we are limited by a lack of understanding by the disease pathology(55). However, the reason most commonly cited as a cause of the "market failure" to encourage NTD research and product development, is their association with poverty. Patient populations lack purchasing power, thus, there is little or no financial incentive for big pharmaceutical companies, the traditional developers of new drugs and vaccines, to create new treatments(50,52,56,57). Consequently, large Pharmaceutical corporations have shied away from NTD product development, with research into new treatments for NTDs once coming to a virtual standstill(50). For certain NTDs, this has prepared the way for a possible health crisis, as resistance to old drugs emerges and disease prevalence increases.

The need for strong policy decisions and national and international governance to support the development of new NTD products has been argued for by several proponents(56,57). Unfortunately, for a time, programs to control/eliminate NTDs were deficient, and NTDs were not formally recognised by health organisations – the WHO originally failed to specifically mention several NTDs in their Millennium Development Goals(38). In recent years, thanks to the restless awareness-raising activities of various researchers and NGOs, the fight against neglected diseases is now supported by clear policies in a number of institutions, globally. This change has been accompanied by unprecedented structural changes in the pharmaceutical industry which have helped to spur the development of a number of new drugs and vaccines for NTDs(57,58).

Recent advances

Some of the policies designed to spur NTD product R&D can be categorised as "push" and "pull" mechanisms. "Push" mechanisms are those that aim to encourage or enable R&D by cutting the costs of doing so. Product researchers and developers usually gain access to these through collaborations with governmental organisations and charities. "Pull" mechanisms encourage investment in R&D by creating financial incentives for this activity. Examples of these are Advanced Market/Purchase Commitments (AMCs) and the US Food & Drug Administration's (FDA) Priority Review Vouchers (PRV)(59,60). Although the effectiveness of these "pull" mechanisms is yet to be proven(59,61–64), they do theoretically solve the problem of a deficient market to incentivize NTD product R&D. The "push" mechanisms have already proven their worth in spurring several new R&D projects, whilst also re-engaging pharmaceutical companies(52,57). The majority of this has happened against a backdrop of a

paradigm shift in pharmaceutical development for NTDs, the product development partnership (PDP)(57,65), which will be discussed in detail below.

Although still much smaller than for other therapeutic areas, the global investment in pharmaceutical R&D for NTDs has increased vastly. In 1999, it was estimated that less than US$70 million were invested into drug research for malaria, tuberculosis, HAT and leishmaniasis, combined. In contrast, the 2012 Global Funding of Innovation for Neglected Diseases Report (G-FINDER) stated that US$ 3,045 million were invested globally in NTD research, with around 70% of that directed towards product development (as reported by respondents to the G-FINDER survey)(66). Although funding, particularly from philanthropic and public donors, has declined since the global financial crisis, funding levels are still higher than in 2007(66), and much greater than 15 years ago(50).

Many national governments are now some of the biggest funders of NTD R&D. Most prominent are the US, through the US Agency for International Development (USAID), and the UK, through its Department for International Development (DFID). Other countries, such as The Netherlands and Australia, have emerged as major NTD R&D funders(66). Meanwhile, the WHO's Special Program for Research and Training in Tropical Diseases (WHO TDR) continues its work with a wide range of private and public R&D partners around the world(66–68).

1.2.3 The Product Development Partnership (PDP) model

PDPs bring together several different partners with the goal of developing one or a related group of products. PDPs are typically partnerships between public and private institutions, where the relative strengths of each individual institution can be exploited. This is particularly useful in developing products for neglected tropical diseases, for a number of reasons. Philanthropic organisations are able to allocate resources to the researchers and developers, covering or reducing the costs of R&D. Academic and other public-sector or non-profit research institutions (PSNRIs) are very-well placed to carry out discovery and early-stage development work(69–71), whilst biotechnology and pharmaceutical firms are well technologically resourced. Large pharmaceutical firms have a wealth of experience in conducting clinical trials, but this can also be supplemented by the knowledge and experience of institutions and NGOs operating in NTD-endemic countries. This ability to leverage the strengths of each partner has likely contributed to the success of the PDP model. It has resulted in clinical development times that are comparable with more profitable pharmaceuticals. Also, average development costs are significantly cheaper than for the development of more profitable drugs, outside of the PDP model(57,72).

Multiple Diseases

- Drugs for Neglected Diseases Initiative (DNDi)*
- European Vaccine Initiative (EVI)
- Foundation for Innovative New Diagnostics (FIND)*
- Infectious Disease Research Institute (IDRI)
- Innovative Vector Control Consortium (IVCC)
- Institute for One World Health (IOWH, a division of PATH)
- International Vaccine Institute (IVI)
- Program for Appropriate Technology in Health (PATH)
- Sabin Vaccine Institute
- World Health Organisation: Special Programme for Research and Training in Tropical Diseases (WHO/TDR)*

Disease-focused

- Aeras Global TB Vaccine Foundation (Aeras)
- Dengue Vaccine Initiative (DVI)
- Global Alliance for TB Drug Development (TB Alliance)*
- International AIDS Vaccine Initiative (IAVI)
- International Partnership for Microbicides (IPM)
- Paediatric Dengue Vaccine Initiative (PDVI, division of IVI)
- Malaria Vaccine Initiative (MVI, a division of PATH)
- Medicines for Malaria Venture (MMV)
- Meningitis Vaccine Project (MVP, a division of PATH)
- Rotavirus Vaccine Program (RVP)
- TuBerculosis Vaccine Initiative (TBVI)

Figure 1.2.3 - examples of PDP "Brokers" involved in product development for drugs, vaccines, diagnostics and insecticides for NTDs. * indicates that they participate in the development of products to control trypanosomiasis(72–74).

Several of the PDPs are co-ordinated by “Brokers”, such as the Drugs for Neglected Diseases Initiative (DNDi) and BIO Ventures for Global Health (BVGH). These partnership brokers are important development partners, due to their extended networks, experience and central role in allocating funding. Disease-specific alliances (e.g. The Medicines for Malaria Venture, MMV) have also emerged as important brokers of PDPs and as developers in their own right.

1.2.4 Imbalanced R&D priorities

Despite these positive changes, a number of concerns in NTD product development still remain. One is the gaps in research, which stem from the lack of sustainable funding for some diseases. This problem will be minimised if global funding increases, and if partnerships continue to be formed appropriately and managed effectively(57). What may potentially be a bigger problem, from a global health perspective and from a researcher’s perspective, is the bias of product development for re-formulations and fixed dose combinations, rather than for new chemical entities(52)(see Table 4.4.2). Trouiller and colleagues describe how the lack of appropriate “tools” to treat certain NTDs, requires that innovative new products are developed – new chemical entities (NCE)s and novel vaccines. However, for drugs in general, and to a greater

extent for NTD drugs, very few NCEs are developed when compared to other registered products(51,52).

1.2.5 Trypanosomiasis

The two human diseases caused by trypanosome infections, human African trypanosomiasis (HAT) and Chagas Disease, are classic examples of NTDs. They are debilitating and fatal diseases that primarily affect the poor. HAT, which is prevalent in Sub-Saharan Africa, and Chagas Disease, which is concentrated in Latin America, both have a long history. Recent efforts have led to improved control of both disease, but problems persist.

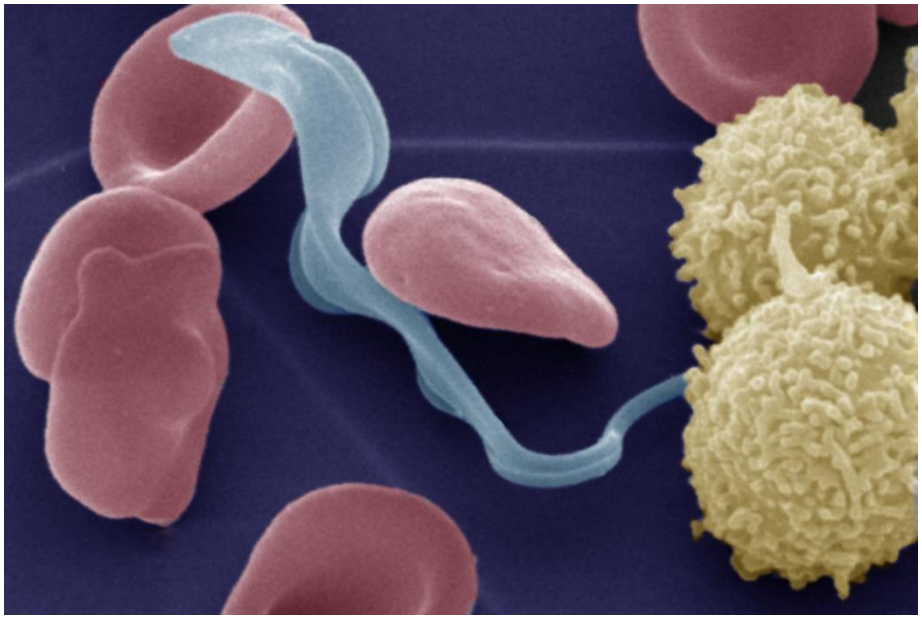


Figure 1.2.4 - coloured electron micrograph of bloodstream form *Trypanosoma brucei* in mammalian blood. Image courtesy of Michael Duszenko, University of Tubingen.

Trypanosomiasis is a NTD caused by infection with species of the genus *Trypanosoma*, a parasitic protozoan distinguished by a single flagellum. This is the cause of two NTDs: human African trypanosomiasis (HAT), which is prevalent in sub-Saharan Africa, and Chagas disease, which is prevalent in Latin America(53,75). HAT, also known as African Sleeping Sickness, is caused by the species *Trypanosoma brucei*, of which there are two main subspecies that cause disease in humans: *T. brucei gambiense* and *T. brucei rhodesiense*. Other species of *Trypanosoma*, including *T. brucei brucei* and *T. congolense*, cause trypanosomiasis in animals(75–77).

T. brucei is spread by vectors of the *Glossina* species, commonly known as Tsetse flies, and the parasite is spread to mammals through the saliva of the vector as it feeds on their blood. Trypanosomes live in the mid-gut of Tsetse flies as replicating procyclic trypomastigotes. These trypanosomes transform into epimastigotes and migrate to the salivary glands, where they

multiply further. The epimastigote trypanosomes mature into metacyclic trypomastigotes, and it is these trypanosomes that are transmitted to mammals when the Tsetse fly feeds. Once inside the host, the trypanosomes transform into bloodstream form (BSF) trypomastigotes, which circulate around the body and replicate in the blood, lymph and cerebro-spinal fluid(76,78,79). Inside the host there is an interchange between "slender" BSF trypanosomes and "stumpy" forms, which are non-proliferative, but are pre-adapted for survival in the insect vector environment and so are transmissible(80,81). The life cycle of *T. brucei* is completed when stumpy BSF trypanosomes are ingested by the Tsetse fly during a blood meal.

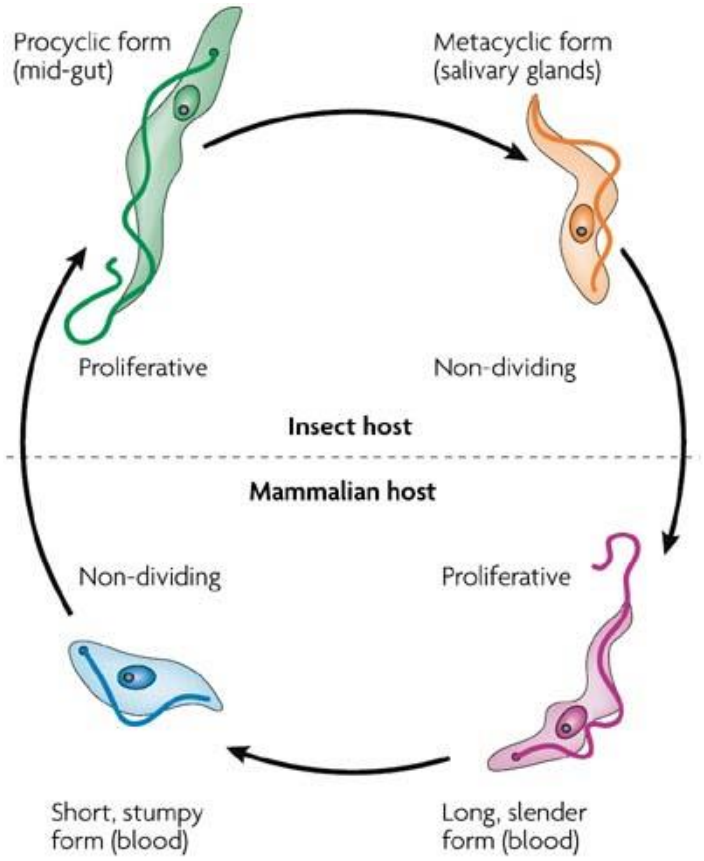


Figure 1.2.5 - summary of the life cycle of *T. brucei*, presented by Lee *et al.*(79).

HAT is characterised by an early stage and a late stage. The disease presents initially with an early haemolymphatic stage, which is characterised by sporadic bouts of fever, and localised swelling at the site of the bite (chancre or Winterbottom's sign). If left untreated, the second, meningoencephalopathic, stage of HAT occurs when trypanosomes enter the central nervous system, leading to a range of psychological and cognitive symptoms, including dyskinesia, tremor, insomnia at night and daytime somnolence, coma and eventually death(76). The pathogenesis of HAT is not completely understood, but symptoms are believed to be mediated largely through inflammatory processes(82). The form of HAT caused by *T. b. gambiense* which accounts for 95-7% of reported cases progresses chronically, and the delay between the haemolymphatic and the meningoencephalopathic stage can be months to years. The *rhodesiense* form of the disease, accounting for 3-5% of cases, progresses much more rapidly,

leading to coma or death within weeks to months if left untreated(54,75,76). Not only is HAT a very debilitating disease, especially in the late stage, but it is also always fatal if left untreated.

Only five drugs are available for HAT treatment, four of which were developed more than fifty years ago(43,44). Pentamidine and suramin are used for early stage *gambiense* and *rhodesiense* HAT, respectively, and melarsoprol and eflornithine are used in the late stage. Melarsoprol is the only drug available for late stage HAT caused by *T. b. rhodesiense*. Nifurtimox has also recently been approved for use in late stage *gambiense* HAT in combination with eflornithine(38,40,43,45,46). The characteristics of these drugs have numerous problems. First of all, pentamidine, suramin, melarsoprol and eflornithine are all administered by injection, which requires well-trained healthcare professionals and appropriate healthcare facilities. This is particularly a disadvantage in the resource-poor settings that HAT is usually encountered. Secondly, several of these drugs are associated with severe side-effects and the hospitalisation and clinical monitoring of patients is recommended. For example, pentamidine injection is painful and there is a significant risk of hypoglycaemia following administration(83). Perhaps the most striking example of adverse effects caused by a trypanosomiasis drug is provided by melarsoprol. The drug, used in late stage HAT, is associated with reactive encephalopathy, a severe side effect of the central nervous system that affects approximately 5-10% of patients and is fatal in around 5% of patients(76,84). For this reason, the corticosteroid prednisolone is co-administered with melarsoprol(85). Diagnosis of HAT is another difficulty as the majority of cases occur in remote rural areas; diagnosis requires detection of symptoms, and confirmation of the presence of trypanosomes in blood and/or cerebro-spinal fluid by lumbar puncture(75,76).

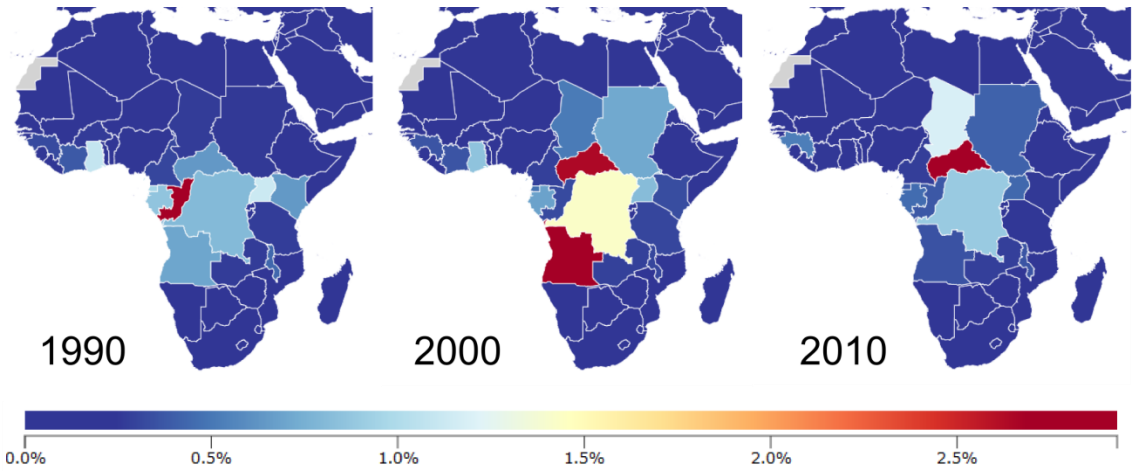


Figure 1.2.6 - the distribution of HAT burden in the years 1990, 2000 and 2010. Countries are coloured according to the percentage of total DALYs caused by HAT in that country. Following an increase in the number of HAT cases and the burden of the disease in the late 20th century, the burden has decreased markedly in several countries. The problem still persists in other countries. Maps were obtained using the IHME GBD Compare tool(47).

HAT is endemic to sub-Saharan Africa and affects mainly rural populations. Approaches to tackle HAT include a combination of vector control, pharmacological treatment and prevention strategies. Despite previous epidemics of HAT in sub-Saharan Africa, vector control and screening campaigns almost eradicated the disease by the mid-1960s. However, a later relaxation of these programs, coinciding with regional conflicts and rising poverty have led to a resurgence in the number of HAT cases. Although these factors still remain barriers to disease eradication, combined efforts of the WHO and various other organisations have helped to decrease the number of reported cases to fewer than 10,000 a year (estimated number of actual cases is 30,000)(75). Animals can be a significant reservoir of *T. brucei* infection, particularly for *T. brucei rhodesiense*, which is usually transmitted from cattle to man. Furthermore, trypanosomiasis in cattle (also called nagana), can affect the livelihoods of farmers in affected areas. Together, the impact of trypanosomiasis on humans and cattle can be a significant impediment to economic progress in the affected areas. Despite the recent advances in the control of HAT, great vigilance is necessary as 70 million people are reported to be living in HAT-endemic areas(54,75).

Chagas Disease, also known as American Trypanosomiasis, is caused by *T. cruzi*, which is spread by insect vectors from the Reduviidae family; vector species include *Triatoma infestans, T. brasiliensis, Panstrongylus megistus, T. sordida* and *T. maculata*(53,86). As humans have encroached on the original habitats of these insects, Reduviid bugs have become adapted to living in domestic environments by feeding on the blood of humans and domesticated animals(86,87). Trypanosomes are present in their replicating epimastigote form in the gut of the vector. They mature into metacyclic trypomastigotes and migrate to the distal end of the intestine. As the bugs feed on mammalian blood they defecate on the skin of the mammal, and *T. cruzi* are able to enter the bloodstream through mucous membranes or through the bite wound made by the bug during feeding. The parasites circulate around the system as trypomastigotes before entering various cells, in which they replicate as the amastigote form. Amastigotes then differentiate into infective cell-derived trypomastigotes, which escape the cells and go on to invade yet more cells. Amastigotes released from cells can also invade macrophages. Transmission from mammalian host to insect vector occurs when a blood meal is taken from an infected mammal(87,88). Other mechanisms for the spread of Chagas Disease include blood transfusions, infected food and congenital infections (maternal-to-foetal transmission)(86).

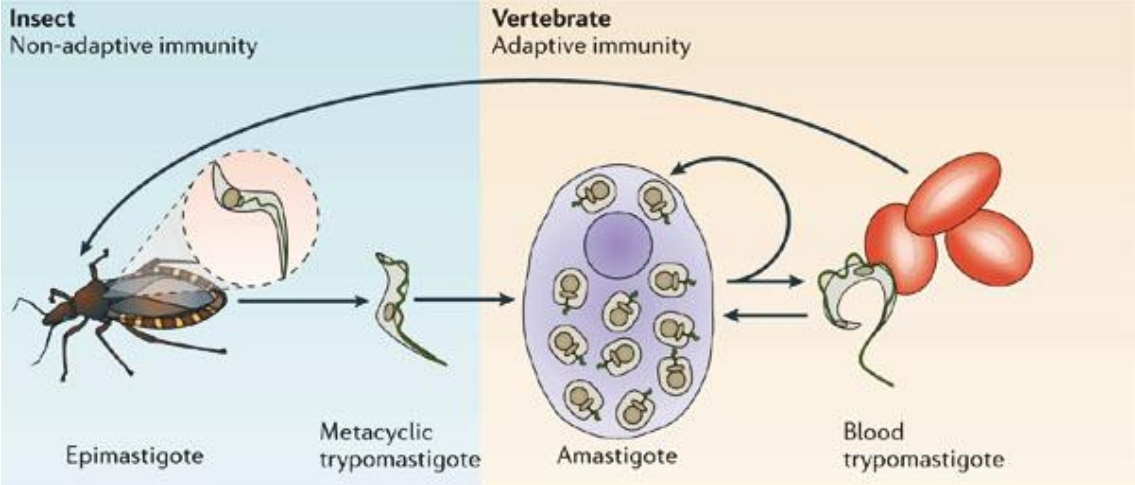


Figure 1.2.7 - the life cycle of *T. cruzi*, presented by Buscaglia *et al.*(88).

Like HAT, Chagas Disease is characterised by an acute phase and a chronic phase. The vast majority of acute infections are asymptomatic. Clinical features of the acute phase include local inflammatory reactions at the site of parasite entry, such as bipalpebral oedema (Romaña's sign) or inoculation chagoma (localised swelling), and other symptoms, including conjunctivitis, fever and general malaise. In more severe cases, organ-specific signs and symptoms may be observed, including, hepatosplenomegaly, myocarditis and meningoencephalitis(86,89). The acute phase is followed by an asymptomatic/indeterminate phase that may last for 5-15 years before clinical signs of the chronic disease appear. The most significant burden of Chagas disease is caused by the symptoms of the chronic disease which are thought to be caused by parasite-triggered autoimmune reactions, which begin in the acute phase and persist for several years, together with the presence of parasites in affected organs. In the most common manifestations of chronic Chagas disease, the heart and the enteric tract are affected, with cardiomyopathy being the complication with the biggest contribution to morbidity and mortality(86,87,89). There are only two clinically approved drugs to treat Chagas disease, benznidazole and nifurtimox. Although both treatments have been shown to be effective in the acute phase of the disease, they have little efficacy in the chronic phase, where supportive treatments are required to deal with the complications. Furthermore, the treatment courses are long, ranging from 1 to 3 months. The use of both drugs are associated with a high incidence of side effects, including headache, dizziness, nausea, vomiting and weight loss(90,91).

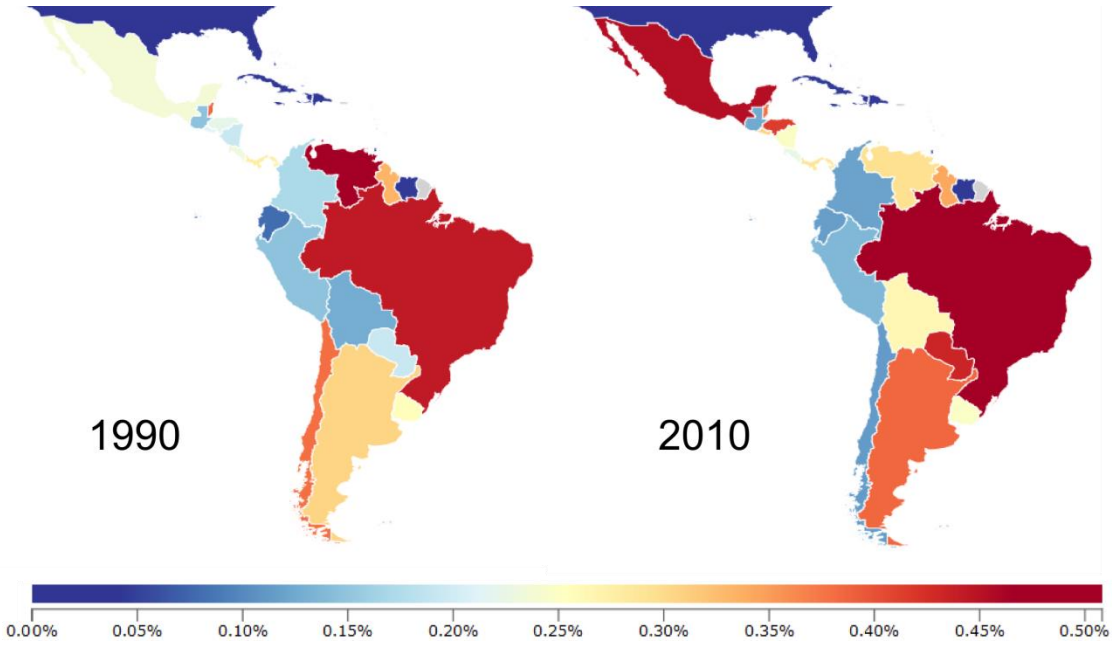


Figure 1.2.8 - the distribution of Chagas disease burden in Latin America in 1990 and 2010. Countries are coloured according to the percentage of total DALYs caused by Chagas disease in that country. The burden of Chagas disease is increasing and spreading to countries further from the original concentration in South America. Maps were obtained using the IHME GBD compare tool(47).

Effective vector control and screening programs in central and south America, where the disease is most prevalent have been very successful in the control of Chagas disease, reducing the annual incidence to around 40 000(53). However, the occurrence of Triatomine resistance to insecticides, the proliferation of new *T. cruzi* serotypes and the export of Chagas disease cases to North America and Europe are creating new risks(87). Up to 10 million people are estimated to be living with Chagas disease, the majority of whom are concentrated in Latin America. However, the incidence has been increasingly rapidly in central and North America, due to migration. There are an estimated 300,000 people living with Chagas disease in the United States and several thousand more cases reported in Canada, Europe, Japan and Australia(92,93). Together, these facts make Chagas disease not only one of the most common NTDs, but also one of the most neglected NTDs in the world(93,94).

Clearly, both human forms of trypanosomiasis represent a significant medical and social need. No one method of control has proven to be totally effective, and the various methods are all limited by certain biological and environmental factors. Furthermore, the physiology of trypanosomes may prevent the development of other tools to control their transmission, such as vaccines. The surface coat of *T. brucei* is composed of a variable surface glycoprotein (VSG), of which there are more than a thousand variants. Only one glycoprotein is expressed at any one time, but the ability to switch between different variants allows these trypanosomes to evade the immune system and for the infection to persist(76,82). For this reason, it is also difficult to develop a vaccine for HAT(95). Few drugs are available to treat HAT and Chagas

disease, and the available drugs cause a high frequency of side effects that are sometimes very severe. The required methods of administration of these drugs, coupled with the lengthy treatment courses and associated costs makes them far from optimal for tackling trypanosomiasis in the relevant settings. As resistance to drugs for trypanosomiasis is increasing, HAT and Chagas disease are good examples of NTDs for which the development of new classes of drugs could provide a significant benefit to health.

Table 1.2.1 - current recommended treatment regimens for HAT and Chagas disease(76,83,90). Treatment regimens vary, so those in this table are representative examples. The recommendations are aimed at the treatment of non-pregnant adults.

Disease	Treatment regimen
HAT	
***T. b. gambiense* early stage**	**Pentamidine**: 4mg/kg intramuscular injection once daily for 7 days
***T. b. rhodesiense* early stage**	**Suramin**: 5 mg/kg intravenously as a test dose, followed by 20 mg/kg on day 3, 10, 17, 24, and 31
***T. b. gambiense* late stage**	**NECT**: **nifurtimox** 5mg/kg orally every 8 hours for 10 days; **eflornithine** 200mg/kg intravenously every 12 hours for 7 days OR **Eflornithine**: 100mg/kg intravenously every 6 hours for 14 days OR **Melarsoprol**: 2.2mg/kg intravenously once daily for 10 days
***T. b. rhodesiense* late stage**	**Melarsoprol**: 2.2mg/kg intravenously once daily for 10 days
Chagas Disease	
Acute stage	**Benznidazole**: 5-7 mg/kg per day orally in 2 divided doses for 60 days
Chronic stage	**Nifurtimox**: 8-10 mg/kg per day orally in 3 or 4 divided doses for 90 days

Desired target product profiles (TPP) for drugs to treat trypanosomiasis have been proposed by a number of researchers(13,96). The TPPs for HAT and Chagas disease are similar, in the respect that an improved side-effect profile and shorter time course of treatment would provide benefits over currently available treatments. Additionally, drugs are sought that would show efficacy against both early and late (or acute and chronic) forms of the diseases. TPPs for a new HAT drug and a new Chagas disease drug are shown in Figure 1.2.9.

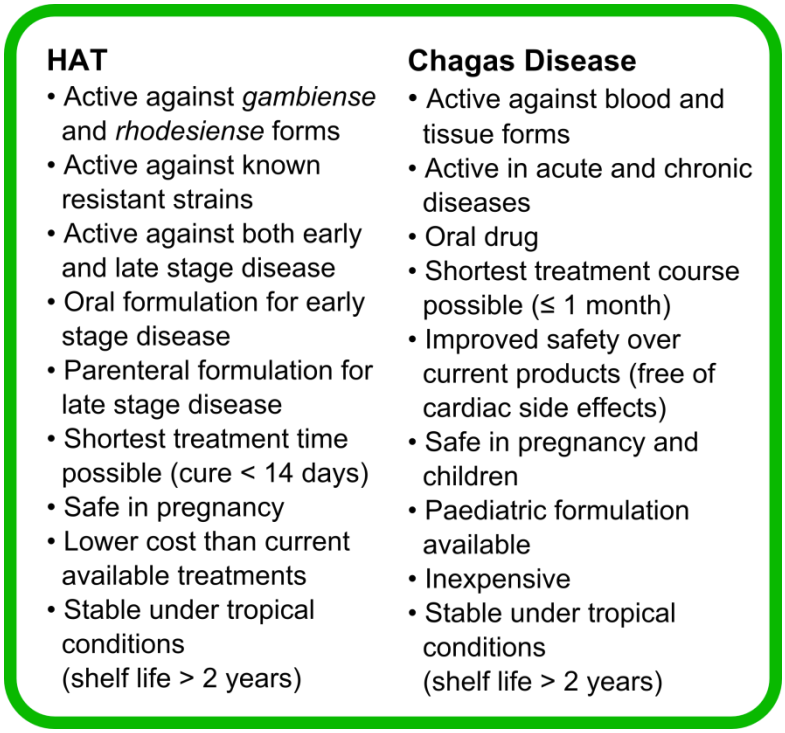


Figure 1.2.9 - a summary of the desired target product profiles (TPP) proposed for new drugs to treat HAT and Chagas disease(13,32,96).

In order to meet some of the criteria set out in the TPPs, an appropriate drug discovery strategy is required. A new drug would need to be potent, but also selective and safe towards the human patient. To achieve this, a rational, target-based approach to drug discovery is particularly effective because differences between host and parasite genomes/proteomes can be exploited. Selecting a particular parasite biomolecule as a target would require that there are significant differences between that target and a human homologue. A more desired property for such a target would be for it to be absent/not utilised in humans. Given the highly unusual physiology of trypanosomes, there are numerous opportunities for target-based drug discovery for antitrypanosomal drugs.

1.2.6 The current status of HAT and Chagas disease control

There has been a significant decrease in the number of new cases of HAT(75). Effective control programs led to a reduction of 68% between 1995 and 2006. 36 countries remain HAT-endemic, although only seven of these have reported more than 100 cases per year in the period 1997-2006(54). However, given the situation in those countries of higher prevalence, along with the ever-present possibility of HAT re-emergence among an at-risk population of 70 million(75), it is important that there are appropriate tools at the disposal of health authorities.

Although control programs (involving screening, vector control and treatment programs) have proven very effective in reducing the number of HAT cases, the effect has been greater for one

form of HAT (the *gambiense* form), because an animal reservoir plays a significant role in the spread and persistence of the other (*rhodesiense)* form(54,75).

Due to the chronic nature of Chagas Disease, diagnosis is challenging and infection often goes unnoticed for several years. Between 7 and 8 million are estimated to be infected with Chagas disease worldwide, and it is spreading beyond its original geographical location to other continents. The treatments are effective if used shortly after infection. However, treatment regimens are long, and adverse reactions occur in up to 40% of patients. The majority of patients will present with chronic infection, often with other complications, when the effectiveness of the usual treatments is limited.

Control of the Chagas Disease-spreading vector, the triatomine bug is probably the most effective means of controlling Chagas Disease, although there are challenges of the re-emergence of the disease in certain areas and the need for more effective treatments to treat chronic infection(53).

1.2.7 Trypanosomiasis R&D

Shortfalls

The biggest shortfall in the control of HAT lies in the tools that are currently available. A HAT vaccine is very challenging, due to the biology of the pathogen(97,98), so efforts are focused on vector control and treatments. All treatment regimens involve some form of injection, which requires more resources and well-trained staff. The length of some of these regimens puts a further strain on resources, which are often not well-developed in rural areas of developing countries. The current drugs all suffer from a high incidence of side-effects, which can be burdensome or even fatal in some cases. Therefore, there is an urgent requirement for new drugs, with less side-effects and more suitable formulations(54,96).

Chagas Disease treatments suffer from similar drawbacks to those of HAT treatments. As stated, above, a new drug for Chagas disease should ideally be more effective than current treatments in chronic stages of the disease and should have a lower incidence of side effects.

1.3 L-threonine 3-dehydrogenase (TDH)

The amino acid L-threonine was first discovered in 1935 following the observation that rats were unable to survive when fed a diet based solely on the 19 known essential amino acids at the time(99,100). Years later, pathways of threonine degradation were elucidated by studying the breakdown of threonine; conversion of threonine to glycine(100–102) and deamination(103) of threonine were some of the earliest processes studied. There are now three known pathways of threonine catabolism, which are initiated by the enzymes L-threonine aldolase (EC 4.1.2.5), L-threonine dehydratase (also known as L-threonine deaminase; EC 4.2.1.16) and L-threonine 3-dehydrogenase (TDH; 1.1.1.103), which is the dominant pathway in several prokaryotes and

eukaryotes(104–106). The activity of L-threonine aldolase has been shown to be low or insignificant in prokaryotes(107) and eukaryotes(108) and no expression of the enzyme has been detected in humans(109). In contrast, the pathway initiated by L-threonine dehydratase is a major pathway for L-threonine catabolism and can become the dominant pathway in certain metabolic states in animals(108). The other major pathway for L-threonine catabolism is initiated by the mitochondrial matrix enzyme TDH, which appears to be the dominant pathway in many prokaryotes and eukaryotes(104–106).

The TDH-initiated pathway, in which L-threonine is converted to glycine and acetyl-coenzyme A (A-CoA), was first described by Elliot(110,111). Later, isolation and characterisation of TDH was achieved by Neuberger and Tait(112,113). Prior to this, aminoacetone had been shown to be produced by the condensation of glycine and A-CoA(114) and the enzyme co-ordinating this was isolated and named aminoacetone synthetase(115). By this time Urata and Granick (1963) had already shown a link between these two reactions(116). It is now known that TDH and aminoacetone synthetase, now more commonly known as 2-amino-3-ketobutyrate ligase (KBL) work in tandem to catabolise L-threonine as illustrated in Figure 1.3.1 below(105):

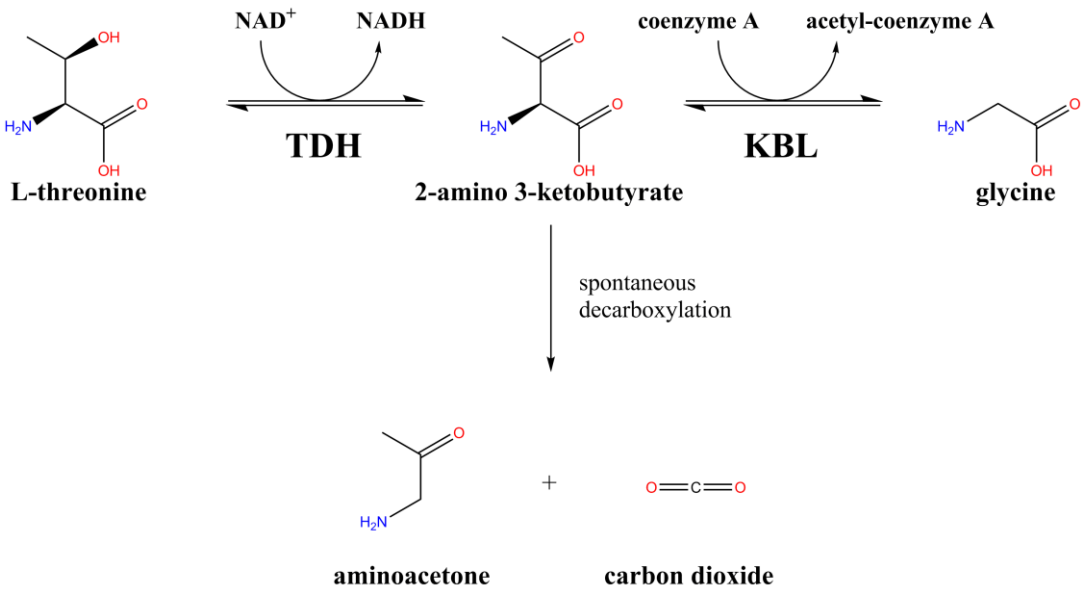


Figure 1.3.1 - the catabolism of L-threonine to glycine and acetyl-CoA. The enzyme TDH oxidises L-threonine to 2-amino-3-ketobutyrate (AKB). KBL then converts AKB to glycine and acetyl-CoA. Alternatively, AKB may undergo spontaneous breakdown to aminoacetone and carbon dioxide.

In the first part of this two-step process, TDH oxidises L-threonine to 2-amino-3-ketobutyrate (AKB) in a reaction dependent on the cofactor nicotinamide adenine dinucleotide (NAD(H)). There is evidence that AKB is unstable and spontaneously breaks down to aminoacetone and carbon dioxide. Thus it has been suggested, amidst other evidence, that TDH and KBL form a multi-enzyme complex to facilitate the completion of the second reaction, the breakdown of AKB to A-CoA and glycine by KBL(104,117)

TDH is a dehydrogenase/reductase (or oxidoreductase). This enzyme family is divided into short-chain (SDR), medium-chain (MDR) and long-chain (LDR) dehydrogenases(118). Several of the most studied TDH enzymes to date have been MDRs, showing a tetrameric quaternary structure and a requirement for divalent metal cations, such as zinc(104,119–123). More recently, a second group of TDH enzymes that share features with the enzyme UDP-galactose 4'-epimerase have been characterised. These enzymes belong to the SDR family, and in contrast to the MDR TDH enzymes, they are dimeric and have no requirement for divalent cations(124–126). The enzyme KBL has not been studied to the same extent, but the crystallographic structure of enzyme from *Escherichia coli* has been published(127).

TDH seems to play important physiological roles in diverse organisms. The TDH pathway has been implicated as an important nutritional source for energy production. A species of *Arthrobacter* was able to grow in medium with L-threonine as its sole 2-carbon source. Metabolised through the TDH pathway, L-threonine was shown to also be a source of nitrogen, and the acetyl coenzyme A (acetyl-CoA) produced by the pathway was fed into the Krebs cycle and glyoxylate cycle for energy production, whilst glycine was thought to be converted to pyruvate(115). Similar observations regarding the role of the TDH pathway were made in *Pseudomonas* spp.(107,128). In several other bacterial organisms, including *Bacillus* sp., *Flavobacterium* sp. and *Corynebacterium* sp., the role of glycine produced from L-threonine by TDH and KBL was highlighted. Glycine produced by the TDH pathway is further metabolised to L-serine or is probably involved in the 'glycine cleavage system' as a source of methylene units(129). In *E. coli*, where L-threonine can act as a source of nitrogen, following TDH production of glycine and conversion to L-serine(130), an additional role for the pathway was identified: aminoacetone, the breakdown product of AKB, was shown to be converted to *D*-1-amino-2-propanol, which is a precursor for vitamin B12 synthesis.

The roles played by the TDH pathway in eukaryotes are just as varied. In rats and pigs, TDH has been shown to play a role in L-threonine homeostasis, and its activity is regulated in response to different metabolic states(108,131,132). Also in rats, L-threonine catabolism by TDH was shown to play a role in fatty acid metabolism through the production of acetyl-CoA. The TDH pathway is also subject to feedback regulation in this role(133). Many of the contributions of the TDH pathway mentioned so far may be transferable between different organisms. However, the pathway has also been implicated in some specialised organism-specific processes. When the plant pathogenic fungus *Sclerotium rolfsii* is grown on L-threonine, there is an increase in branching of the fungus' hyphae and increased formation of sclerotia, dense formations of hardened mycelium that are formed in response to stressful conditions and contain carbohydrate stores. TDH was thought to provide the necessary energy and carbohydrate precursors by supplying acetyl-CoA and glycine to the glyoxylate and glycine-serine pathways(134). Perhaps the most interesting physiological role played by TDH has been discovered in mouse embryonic stem cells (ES cells). The ES cells show a

dependence on L-threonine, and TDH activity is shown to be approximately 100-fold higher in these cells, compared to other mouse cells(135). The dependence is so striking, that exposing those cells to a small molecule inhibitor of TDH was able to selectively kill ES cells. As those cells were perceived to be in a highly energetic state, it was postulated that the glycine produced by the TDH pathway was necessary for the glycine cleavage system(136), as suggested in other organisms(129). In 2012, Shyh-Chang and colleagues reported the fascinating observation that TDH activity correlated with S-adenosyl methionine (SAM) production, which in turn correlated with increased epigenetic regulation by histone methylation. This epigenetic regulation was shown to be essential in maintaining the pluripotency of mouse ES cells(137).

1.3.1 The role of TDH in trypanosomes

As in other organisms in which the TDH pathway has been studied, the pathway was found to play a pivotal role in the physiology of trypanosomes. In 1975, Cross *et al.* reported that cultured *T. brucei* in mixed media preferentially and exhaustively consume L-threonine(138). By using radiolabelled L-threonine, it was determined that the majority of the absorbed L-threonine was metabolised through the TDH pathway, producing acetyl-CoA and glycine. Furthermore, it was discovered that exposure of *T. brucei* in culture to tetraethyl thiuram disulphide (TETD), an inhibitor of human aldehyde dehydrogenase and a potent inhibitor of TDH, caused the rapid death of the trypanosomes. This trypanocidal effect correlated with the inhibition of TDH(138,139). Thus, this represented the first identification of TDH as a potential drug target. TDH is believed to be important in *T. brucei* because the acetyl-CoA produced by the pathway is essential for fatty acid synthesis.

In contrast to other parasites, which obtain their lipids from the host, parasites of the family *Trypanosomatidae* appear to synthesise fatty acids *de novo*(79,140). In *T. brucei*, there are two mechanisms for fatty acid synthesis. The dominant mechanism occurs through a number of endoplasmic reticulum-associated elongase enzymes (ELO), which synthesise and extend fatty acid chains, beginning with malonyl-CoA and butyryl-CoA. The second mechanism occurs within the mitochondria and is mediated by a Type II fatty acid synthase. ELO-mediated fatty acid synthesis has been shown to be essential for the production of myristate, which forms the glycosylphosphatidylinositol (GPI) anchor of the VSG expressed on the surface of BSF *T. brucei*(79). Mitochondrial fatty acid synthesis has been estimated to constitute approximately 10% of total fatty acid synthesis in procyclic *T. brucei* and it is thought to serve a specialised role related to the functioning of the mitochondria itself(79,140,141). Firstly, the fatty acid synthesis is involved in the production of precursors of lipoic acid, a cofactor for pyruvate dehydrogenase (the importance of PDH will be discussed below) and other mitochondrial enzymes(141). Secondly, the inhibition of mitochondrial fatty acid synthesis has been shown to inhibit respiration and effect the morphology of the mitochondrion *in vitro*(140). Thirdly,

compounds that inhibit type II fatty acid synthesis have been shown to inhibit BSF trypanosome growth(79).

The acetyl-CoA required for fatty acid synthesis in *T. brucei* arises from two distinct pools: one produced through the TDH pathway and a second pool produced by the glycolytic pathway, in which acetyl-CoA is produced by catabolism of pyruvate by pyruvate dehydrogenase (PDH)(142–144). TDH, KBL and PDH all reside in the mitochondria, and acetyl-CoA is either used for type II fatty acid synthesis or it is converted to acetate and transported to the cytoplasm for use in fatty acid synthesis by ELOs. Figure 1.3.2 illustrates the production of acetyl-CoA from L-threonine and glucose in *T. brucei*:

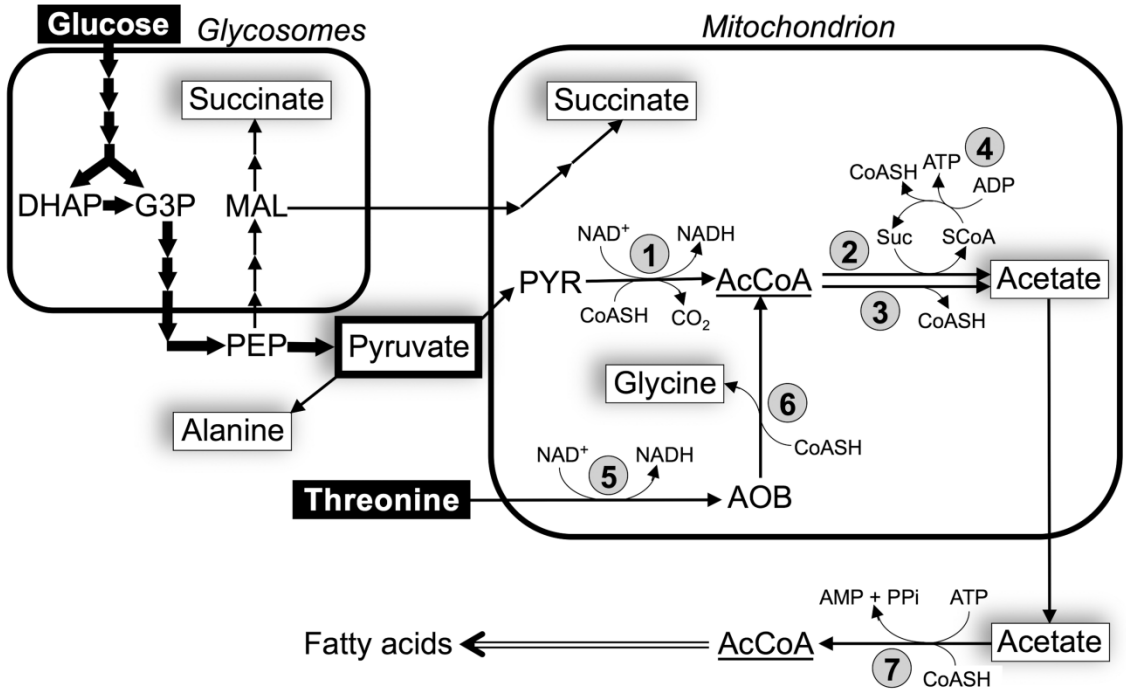


Figure 1.3.2 - the production of acetyl-CoA and acetate in *T. brucei*. Image presented by Mazet *et al.*(144). AcCoA = acetyl-CoA; AOB = amino oxobutyrate (2-amino-3-ketobutyrate, AKB), DHAP = dihydroxyacetone phosphate; G3P = glyceraldehyde 3-phosphate; MAL = malate; PEP = phosphoenolpyruvate; PYR = pyruvate; SCoA = succinyl-CoA; CoASH = coenzyme A. Enzymes are indicated by the following numbers: 1 = pyruvate dehydrogenase (PDH); 2 = acetate:succinate CoA-transferase; 3 = acetyl-CoA thioesterase; 4 = succinyl-CoA synthetase; 5 = L-threonine dehydrogenase (TDH); 6 = 2-amino 3-ketobutyrate ligase (KBL); 7 = acetyl-CoA synthetase.

Although there is evidence that L-threonine is the preferred source of acetyl-CoA for the production of fatty acids in *T. brucei* (142), gene silencing studies have shown that inhibition of TDH alone is not lethal to parasites and does not inhibit growth(143–145). Likewise, inhibition of acetyl-CoA production by PDH on its own does not inhibit trypanosome growth. However, Millerioux *et al.* reported that dual inhibition of TDH and PDH by RNA interference (RNAi) caused powerful inhibition of the growth of procyclic form *T. brucei*. Singular inhibition of TDH or PDH in the absence of glucose or L-threonine, respectively, also inhibited trypanosome growth(143), indicating that concomitant disruption of both routes to acetyl-CoA production is a powerful approach to killing trypanosomes. Despite these results, there was still some doubt

over the effectiveness of inhibiting TDH as part of a strategy to treat trypanosomiasis. The expression of TDH is known to be significantly up-regulated in procyclic *T. brucei* compared to bloodstream forms(139,146–148), apparently as part of the organism's adaptation to the environment of the mid-gut of the Tsetse fly, where more L-threonine is available, but glucose is less abundant. In the study of Millerioux *et al.*(143) and in the studies of *Cross* and colleagues(138,139) that showed TETD to be lethal to trypanosomes, the procyclic form of *T. brucei* was used in growth inhibition studies. However, Mazet *et al.* recently reported that acetyl-CoA production from both glucose and L-threonine is still essential in BSF *T. brucei.* The effects of dual inhibition of both pathways were the same as those observed in the procyclic parasites(144).

There may be yet more ways in which TDH inhibition may be effective for treating trypanosomiasis. Glycine produced from L-threonine by TDH has been shown to be incorporated into trypanothione in *T. brucei*(143). Trypanothione is an essential thiol that is unique to kinetoplastids (includes Trypanosomatidae) and protects them against oxidative stress, heavy metals and toxic metabolites, in much the same way that glutathione does in humans(149,150). The discovery of trypanothione's importance has led to the selection of enzymes involved in its production as drug targets for HAT, Chagas disease and Leishmaniasis, which is a NTD caused by parasites of another Trypanosomatidae genus, *Leishmania*(31,151). Whether the TDH pathway in *Trypanosoma* undertakes some of the functions discussed for other organisms mentioned above remains to be discovered(152,153). Although *Leishmania* parasites lack TDH, the enzyme is present in *T. cruzi*, so TDH could be a suitable target for both African and American forms of trypanosomiasis.

A key consideration in the selection of any target for drug discovery is whether inhibition of that target is likely to cause toxicity to a patient. In this respect, TDH is viewed favourably as a target, as humans do not make use of the enzyme. The TDH gene in humans is an "expressed pseudogene"; the TDH gene contains a number of mutations, which lead to the production of truncated, non-functional proteins when translated(106). The complete absence of TDH activity has also been confirmed in new-born infants(154). The data reported by Alexander *et al.* on the killing of mouse ES cells(136) using a TDH inhibitor is a testament to the effectiveness of TDH inhibition for selectively killing a particular cell type. The mapping of the genomes of *T. brucei* and *T. cruzi* has now been completed, so genetic information on TDH is freely available(152,155). This enables the further study of this important enzyme as a drug target.

1.4 Hypothesis

The aim of the work presented in this thesis was to employ rational methods to the discovery of novel inhibitors of TDH. The results of this effort should provide information to enable the design of effective leads for the treatment of trypanosomiasis. Presented herein is a description of the structural and functional characterisation of TDH using a combination of X-ray crystallography, enzymatic assays and other biochemical methods. The structural information

on TDH was used in a virtual screen of several thousand compounds to identify potential inhibitors of TDH. Subsequently, some compounds identified by virtual screening were assayed alongside other compounds from commercial compound libraries in a medium-throughput *in vitro* screen. The characteristics of these inhibitors will also be discussed in detail. Finally, in view of the commercial and practical considerations necessary for advancing an NTD drug to market, a discussion of the various ways in which NTD products are advanced to the clinic is included.

2. Methods

2.1 Structural Studies of TDH

2.1.1 Protein Expression and Purification

2.1.1.1 Transformation and Expression

Prior to this work, a gene of L-threonine 3-dehydrogenase (TDH) had been amplified using standard high-fidelity PCR methods from genomic DNA from *Trypanosoma brucei* strain TREU 427 and was kindly donated by Professor Jon Kelly from the London School of Hygiene and Tropical Medicine. This DNA was introduced into a pET-15b vector (Novagen, Darmstadt).

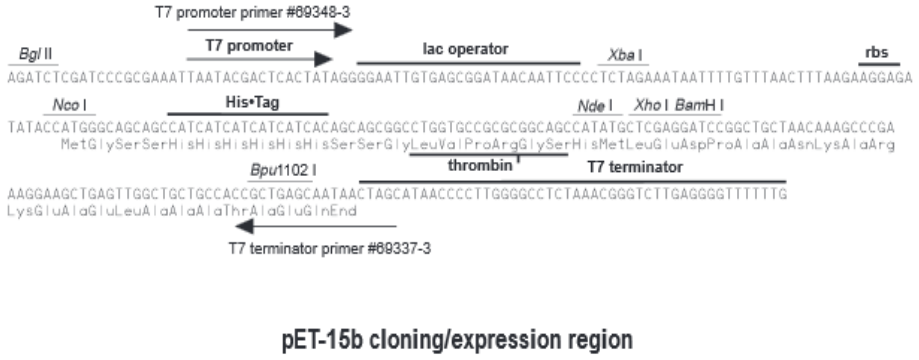


Figure 2.1.1 - p-ET15b vector with multiple cloning sites. Highlighted sequences of relevance include the lac operator (allows induction of expression by lactose analogues), and the His-tag (a sequence of six histidine residues, useful for purification).

Escherichia coli BL21 (DE3) or Rosetta® cells (Novagen, Germany) were used as competent cells for the expression of protein. Competent cells were transformed with the TDH DNA by introduction of a 45ng.ml^{-1} solution of DNA into a 50-100 µl suspension of competent cells. The cells were then stored in an ice bath for ten minutes and it was periodically flicked to disperse the DNA. These competent cells are stored without nutrients in a solution of calcium chloride to make them more permeable to DNA. To further encourage the uptake of DNA, the cells were given a heat shock for 45 seconds at 42°C in a water bath, before being transferred to an ice bath for another ten minutes. After this, they were transferred to room temperature and 1ml of autoclaved Luria Broth (LB) was added. The cell population was grown by incubating in a rotating water bath at 37°C and 90 revolutions per minute (rpm) for 25 minutes. This culture was then spun at 13,000rpm in a microcentrifuge for 1 minute. This speed may be higher than that used in many labs, but it was suitable for the purposes of this procedure. Excess broth was removed, and the remaining cells were re-suspended by trituration using a micropipette. This suspension was pipetted onto a previously made agar plate and spread across the surface using a glass spreader. The agar plate was incubated overnight at 37°C. The plate was then

stored at 2-8°C until ready for use. Four separate cultures of these bacteria were prepared by inoculating 10ml LB in bijou bottles, using four isolated colonies picked from the agar plate. The cultures were grown in the presence of 50 or 100µg/ml ampicillin to prevent contamination with other bacteria and to ensure that the expression vector was retained. The cultures were grown overnight in a shaking incubator at 37°C and 120rpm. These cultures were to serve as stock cultures of TDH-expressing cells, so a test of their ability to express TDH from the introduced DNA was carried out. Four 10ml cultures in bijou bottles were prepared by inoculating LB plus ampicillin with 300µl of the stock cultures. These cells were incubated at 37°C and 120rpm until they reached the mid logarithmic phase (optical density at 600nm [OD_{600}] of 0.4-0.6). Each culture was divided in two, leaving eight 5ml cultures, four of which would be experimental cultures and the other four would be controls. The experimental cultures were all induced to express TDH by the addition of 1mM of the lactose analogue, isopropyl-β-D-1-thiogalactopyranoside (IPTG). All eight cultures were grown overnight in an Innova43 shaking incubator at 37°C and 120rpm as previously. The IPTG-induced expression of TDH was detected using SDS-polyacrylamide gel electrophoresis (SDS-PAGE). SDS-PAGE is able to detect and differentiate between proteins of different charge and mass, based on their differential migration through a polyacrylamide gel over time on application of an electrical current. This migration is therefore governed by the polyacrylamide content of the gel, the electrical current and the mass-to-charge ratio of the protein. Samples were prepared by centrifuging 300µl aliquots of each culture at 13,000rpm for 5 minutes, discarding the supernatant and re-suspending the cell mass by trituration in 50µl of Laemmli buffer. Laemmli buffer contains a number of constituents which are very important for the method to work. Firstly, the die, bromophenol blue, allows researchers to visually track the movement of proteins through the gel. Most importantly, another constituent, sodium dodecyl sulphate acts as a protein denaturant and surfactant that causes the unfolding of proteins and coats them in a negative charge, allowing them to be moved by electrophoresis. Samples were boiled for 10 minutes, to ensure protein denaturation, prior to loading into wells in 12% Tris-glycine SDS-PAGE gels. SDS-PAGE was performed by applying an electric current across the gel at 30milliamps in a Bio-Rad Mini-PROTEAN II electrophoresis cell, powered by a Bio-Rad Power Pac 300 (Bio-Rad Laboratories Ltd., Hertfordshire, UK). Protein molecular weight markers (Fermentas, part of Thermo Fisher Scientific Inc., Leicestershire, UK) were also loaded onto gels in a separate well. Gels were stained with Coomassie blue and de-stained with a solution consisting of 10% v/v acetic acid, 10% v/v methanol and 80% v/v distilled water. Successful expression was confirmed by the presence of a strong blue band corresponding to a molecular weight around 38kDa in the induced cultures. Expression of KBL was performed using the same protocol. KBL expression was confirmed by SDS-PAGE with the presence of a band at around 46kDa.

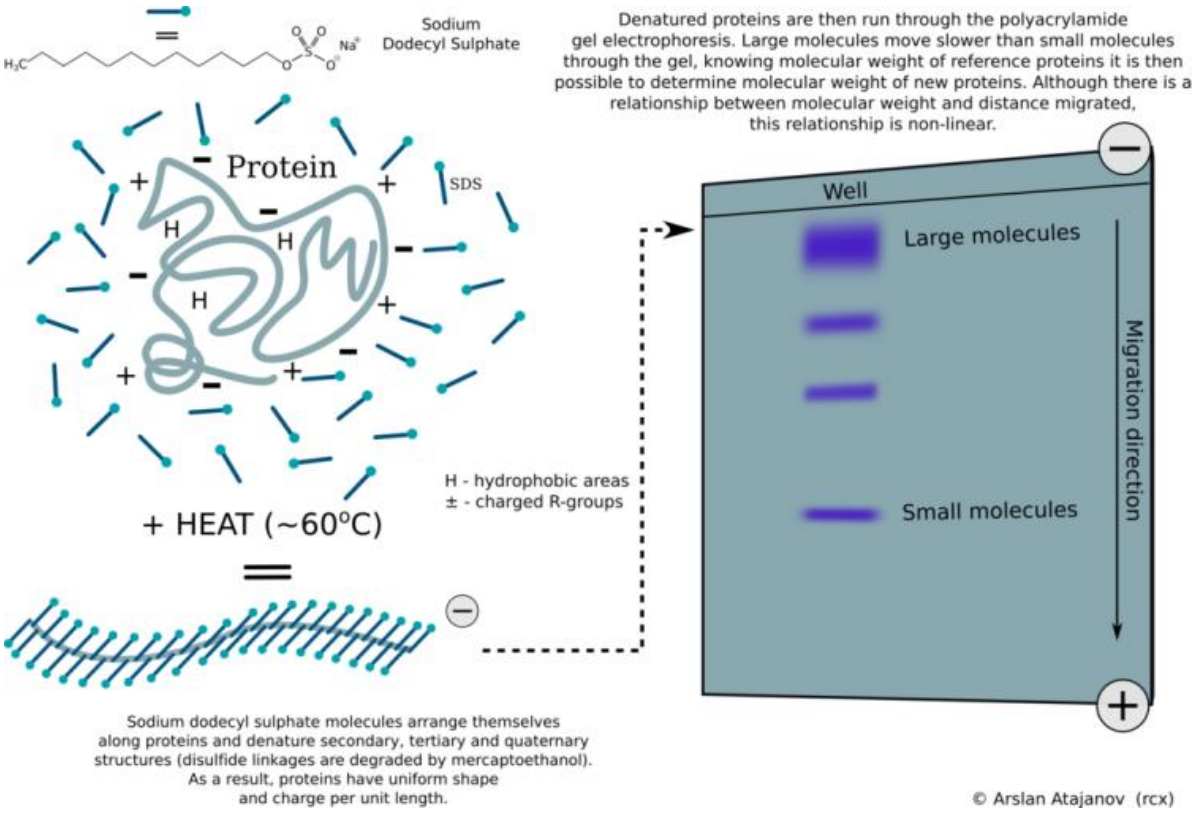


Figure 2.1.2 - schematic summarising the use of SDS-PAGE to analyse proteins. Figure obtained from (156).

2.1.1.2 Expression and Purification

1 litre broth cultures for protein expression were prepared in 2-litre conical flasks by inoculating LB plus ampicillin with 10ml of TDH or KBL expression-inducible stock culture. These cultures were then grown by either of the following methods:

- 'Normal' method: Cultures were grown to the mid logarithmic phase and induced with 0.3mM IPTG. Cells were then incubated overnight at 37°C in a shaking incubator at 120rpm.
- 'Heat-shock' method: Cultures were grown to a level between the early log and mid log phases (OD_{600} 0.3-0.4) before being transferred to a rotating water bath, where they were given a heat-shock, by incubating at 42°C whilst rotating at 90-100rpm for 30 minutes. Exposing the bacteria to heat stress leads to the increased expression of a number of heat shock proteins, which are molecular chaperones. These proteins bind to hydrophobic regions of proteins, thereby assisting in the folding process and reducing the degree of misfolding and denaturation that often occurs during induced expression of high levels of recombinant protein(157,158). After the heat shock, the cultures were cooled by swirling in an ice bath for 5-10 minutes before being induced with 0.3mM IPTG. Cells were then incubated at 18°C in a shaking incubator at 90rpm for 48-72 hours.

Subsequently, cultures were transferred to appropriate containers and were centrifuged for 25 minutes at 7000rpm in an Avanti J-26 XP centrifuge (Beckman Coulter, California, USA) with a JLA 8.100 rotor. The supernatant was discarded and the cell mass re-suspended in 'lysis' buffer (50mM Tris.HCl, 100mM sodium chloride, 1mM β-mercaptoethanol (βM) and 10% v/v

glycerol, pH 7.5). This suspension was then sonicated using a Soniprep 150 sonicator (MSE, London, UK) with an exponential probe, vibrating at an amplitude of 10 microns to lyse the cells. Fifteen 30-second cycles of sonication, separated by 30 second intervals were applied, before mixing the cells by stirring or shaking and then repeating the process once more. The cell lysates were then poured into Oak Ridge tubes, and centrifuged at 18000rpm at 4°C for 40 minutes using a JA 25.50 rotor (Beckman Coulter, California, USA). Both the resulting supernatant and cell pellet were retained and samples were tested by SDS-PAGE, as described above.

Once TDH was confirmed to be in the supernatant, it was purified by affinity chromatography using a HisTrap® HP nickel column (GE Healthcare). The rationale behind this method is that the protein is 'tagged' with a tail of six histidine residues, which bind strongly to Ni^{2+} ions in the column, whilst other proteins flow through the column. The purified protein can then be recovered by addition of a solution containing imidazole, which is able to out-compete histidine for binding. The purification process was carried out using the following buffers:

- Binding buffer: 50mM NaH_2PO_4, 300mM NaCl, 10mM imidazole, 1mM βM, pH8.0;
- Wash buffer: 50mM NaH_2PO_4, 300mM NaCl, 20mM imidazole, 1mM βM, pH8.0;
- Elution buffer: 50mM NaH_2PO_4, 300mM NaCl, 250-500mM imidazole, 1mM βM, pH8.0.

The binding and wash buffers contain low concentrations of imidazole to prevent the non-specific binding of other proteins, whilst the elution buffer uses high imidazole concentrations to elute the protein of interest. Buffers and the supernatant were introduced into the column through a 0.45micron filter using a 10ml syringe or a peristaltic pump and a protocol similar to the manufacturer's instructions was used. They were injected into the column in the following order: 5ml of elution buffer to clean the column, 10ml of binding buffer to equilibrate the column, the entire supernatant (containing protein), 10-15ml of wash buffer to remove non-specifically binding substances and finally 5ml of elution buffer was used to elute the protein from the column into 15ml of lysis buffer to maintain solubility of the protein. According to the manufacturer, the HisTrap® HP column has a protein binding capacity of 40mg per ml of resin. To reduce the loss of unbound protein when this capacity was greatly exceeded, two HisTrap® HP columns were attached end to end. The procedure was then carried out in the same way, but using double the volumes of each buffer. All solutions flowing through the column(s) ("flow-throughs") were collected for SDS-PAGE analysis, where the success of the procedure was confirmed. The eluted protein was concentrated and washed with around 40ml of lysis buffer using ultrafiltration in a 20ml, 10,000MW cut-off Vivaspin® 20 tube (GE Healthcare) to remove imidazole and to concentrate the protein. The tube was spun in an Allegra X-12R centrifuge (Beckman Coulter) at 3000-3500rpm and at 4°C. Centrifugation was continued until TDH was concentrated to a value of 1-20mg/ml. Finally, the filtrate and retentate were tested by SDS-PAGE as described above to confirm TDH purity. The same procedure as that outlined above was used to purify KBL, except the heat-shock expression method was always used and the lysis buffer contained 300mM NaCl instead of 100mM NaCl.

In some cases, the purity was judged by SDS-PAGE to be sufficient for further studies. The 'storage buffer' used was the same as the 'lysis buffer': 50mM Tris.HCl, 10% v/v glycerol, 1mM βM, and 100mM or 300mM NaCl to store TDH or KBL, respectively.

In circumstances where it was desirable to ensure higher purity (i.e. before the *in vitro* screening experiments), or when the enzyme solution still contained significant impurities, ion exchange chromatography was used. Ion exchange chromatography works on the principle that charged molecules will bind to a polymer resin, and then those charged molecules can be displaced by the addition of ions. Different molecules will either not bind or they will bind to the resin and later be displaced by increasing ion concentrations, thus allowing separation of a desired protein from impurities. For this technique to work, the buffer used on the column, must have a pH $\geq$ 1 unit above or below the relevant protein's isoelectric point (pI, the pH value where a molecule is neutral/carries no charge) to ensure that the protein is positively or negatively charged. Amino acid sequences for TDH and KBL were obtained from the TriTryp online database(155) and the pI values were predicted using the ProtParam tool on the ExPASy server(159). Ion exchange chromatography was carried out on a column of approximately 250ml Q Sepharose resin. Q sepharose is a cross-linked polysaccharide resin with positively charged quaternary ammonium ionic groups that bind negatively charged molecules. Accordingly, the proteins were dissolved in buffers with pH values above their pIs, so that they carried a negative charge. Prior to use, the column was washed of any bound substances using 1000ml of a 'high salt' buffer, containing 50mM Tris.HCl, 10% v/v glycerol and 1M NaCl. The column was then equilibrated with 1000ml of a 'low salt' buffer, containing 50mM Tris.HCl, 10% v/v glycerol and 100mM NaCl. The protein was then injected onto the column. The column was washed with low salt buffer until some protein was eluted from the column, as detected by UV absorbance at 280nm. This protein was collected and retained. The procedure was continued by running a gradient of 1000ml solvent, progressing from 100% of the low salt buffer to 100% high salt buffer. Ion exchange chromatography was automated by an ÄKTA Prime (GE Healthcare) Fast Performance Liquid Chromatography (FPLC) machine. Eluted proteins were collected in fractions and concentrated and washed in storage buffer by ultracentrifugation, as described above. The identities of proteins in the unbound and bound fractions were determined by SDS-PAGE.

TDH and KBL solutions were stored in ice baths and refrigerated at 2-8°C. Alternatively, solutions were stored for longer periods, frozen at -20°C.

2.1.2 X-ray Crystallography and Structure Analysis

X-ray crystallography is used to model the three-dimensional structure of small molecules and macromolecules, using the patterns formed by the diffraction of X-rays by crystallised substances. From these diffraction patterns, the three-dimensional structure of the diffracting structure can be calculated by use of a Fourier transform. The use of X-ray crystallography to

determine the structure of proteins, as used to determine the structure of TDH and KBL, will be outlined below.

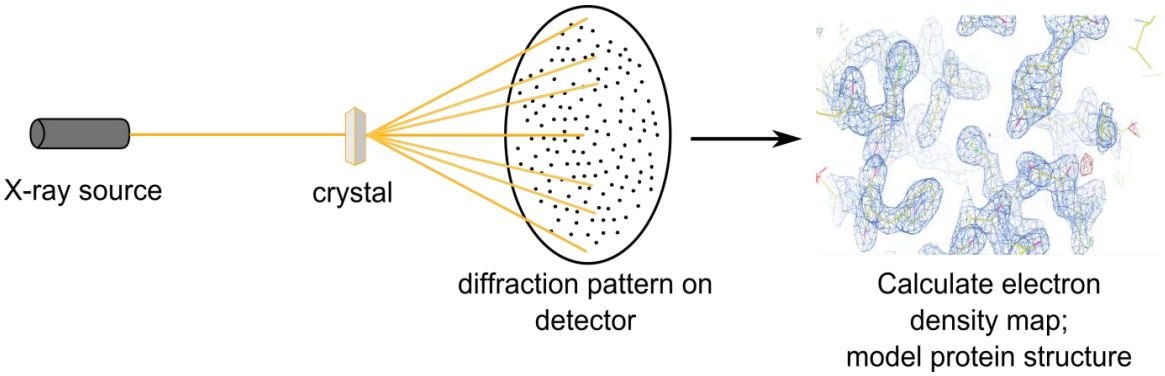


Figure 2.1.3 - X-ray crystallography schematic. An X-ray beam is diffracted by a crystal. Secondary X-rays are recorded on a detector. The diffraction pattern can be used to calculate an electron density map of the diffracting substance.

2.1.3 Crystallisation

Proteins were crystallised by mixing solutions of proteins with crystallisation solutions. 3-5µl of protein solution was added to the same volume of precipitant solution to form droplets of 6-10µl. Crystallisation solutions contain precipitants, which promote the formation of crystals through a number of mechanisms, particularly by decreasing the solubility of the protein. Molecular crowding agents, such as polyethylene glycol also help to encourage crystallisation by forcing different protein molecules into contact with each other. Droplets of protein and precipitant are stored in a sealed container which contains a reservoir of crystallisation solution. In the sealed environment, crystals may be formed by vapour diffusion: as water evaporates from the reservoir and the protein-precipitant solution, the concentrations of both protein and precipitant increase, promoting crystallisation. Crystallisation of TDH and KBL was attempted by the 'hanging drop' method (see Figure 2.1.4), where the droplet of protein and precipitant solution is suspended from a siliconised glass plate (cover slip).

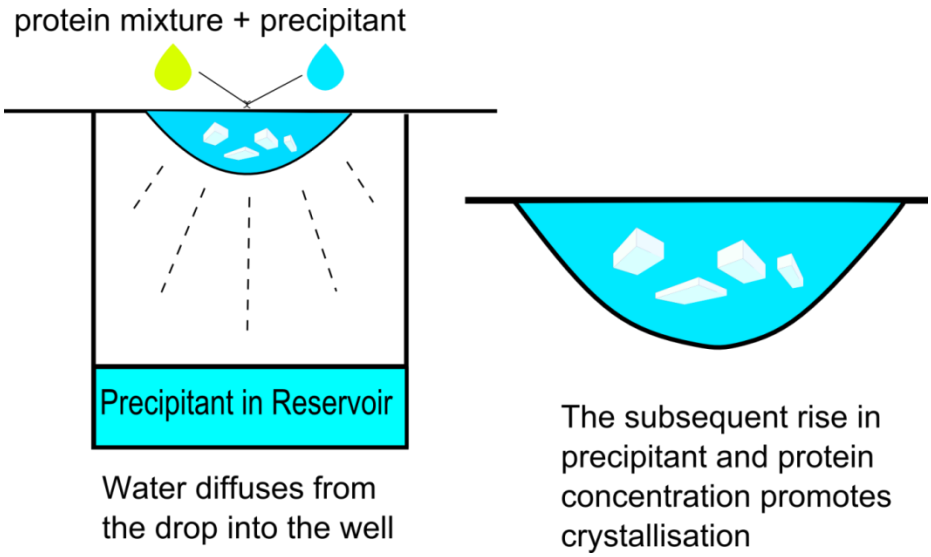


Figure 2.1.4 - Crystallisation by the hanging drop vapour diffusion method.

The crystallisation of any particular protein can be affected by numerous factors, including the protein concentration, precipitant identity, precipitant concentration, pH, salt identity and salt concentration. For this reason, crystallisation was attempted by combining different concentrations of protein with commercially available screens, which contain an array of crystallisation conditions that combine different precipitants, salts, buffers and pH values. 'Structure Screens 1 and 2' (MD1-01 and MD1-02, Molecular Dimensions) and 'JCSG *plus*' (Molecular Dimensions) screens were used to identify conditions that produced crystals of TDH. Several TDH crystals produced from these screens were later used to collect diffraction data. The successful crystallisation conditions were reproduced using solutions prepared using the reported constituents. With the aim of improving the quality of TDH crystals, some conditions identified in the commercial screens were altered by use of an optimisation matrix. This involves combining different concentrations of precipitant with different pH values. A similar process was used to promote crystallisation of KBL in conditions that produced amorphous solids, or crystalline-like solids.

An important utility of X-ray crystallography in a drug discovery context is the ability to observe ligand-protein interactions. Therefore, in order to obtain diffraction data from a protein interacting with its ligand, the ligand was introduced by one of two methods. Using the method of co-crystallisation, the ligand is mixed with the protein prior to crystallisation. Ligand solutions were made up in the same storage buffer as the protein or in DMSO if aqueous solubility of the ligand was low. The ligand solution was then mixed with the protein solution and left for 30 minutes to an hour before carrying out crystallisation as described above. A second method, often called 'crystal soaking', involved transferring crystals to a droplet of crystallisation solution, which also contained ligand. In this way, crystals were soaked in a solution containing the ligand for 5 – 60 minutes.

X-ray radiation damages biological materials, thus making it difficult to collect reliable or complete diffraction data. To limit this damage, crystals are 'cryo-cooled' to low temperatures of around 100K and maintained at these temperatures throughout storage and data collection. It is important that cooling happens rapidly and that crystals are not frozen, which would cause changes in the physical structure of the crystal. With this in mind, after 'harvesting' protein crystals onto a fabric or plastic loop, they were cooled by one of three methods:

- The loop and crystal were plunged directly into a bath of liquid nitrogen;
- The loop and crystal were suddenly exposed to a stream of cryo-cooled nitrogen gas;
- The loop and crystal were plunged into liquid ethane for one second and then into liquid nitrogen.

To prevent damage to the crystals on cooling, the crystals were sometimes soaked in glycerol, which acts as a cryoprotectant. Crystals were transferred to a drop of 10µl crystallisation solution, and 4 x 1µl drops of glycerol were added and mixed into this drop sequentially. Crystals were then harvested and cooled. All cooled crystals were stored in a Dewar to maintain their temperature, prior to data collection.

2.1.3.1 X-ray Diffraction

The determination of macromolecular structures from X-ray diffraction patterns requires knowledge of the properties of waves. X-rays are transverse waves and their properties are outlined in Figure 2.1.5. The wavelengths of X-rays range between 0.1 and 100 Å, with the useful range for crystallography lying between 0.7 and 1.5 angstroms, as they match interatomic bond lengths (the typical carbon-carbon bond is 1.5 Å in length).

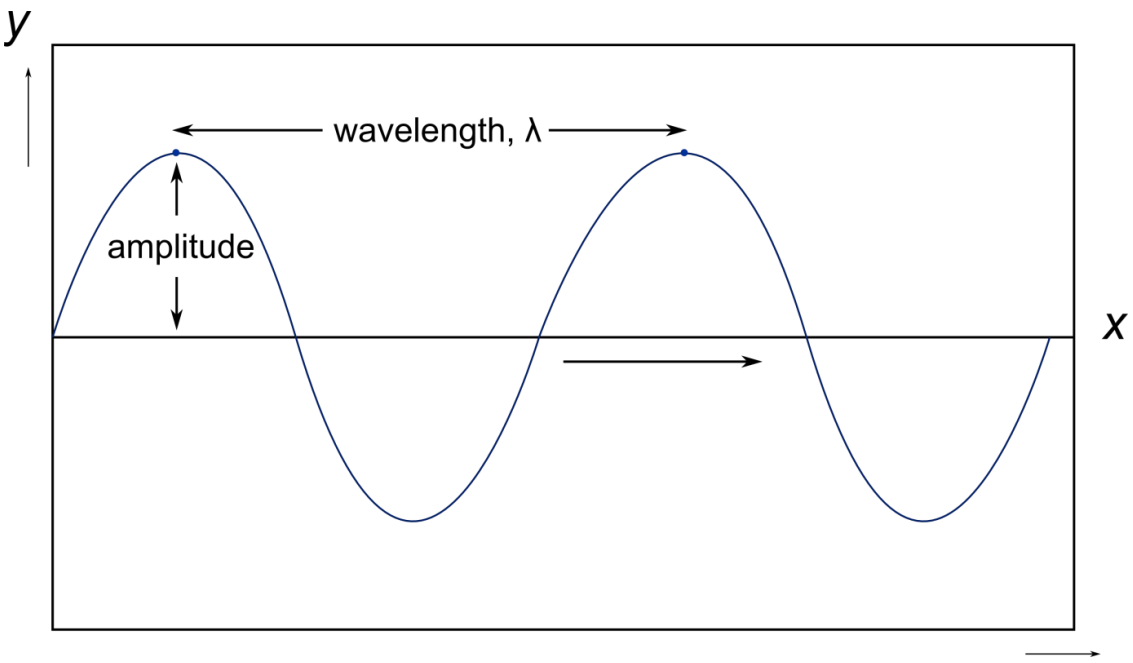


Figure 2.1.5 - the properties of a transverse wave (shown in blue). The sinusoidally oscillating wave is described by its amplitude, the distance from its equilibrium point to a wave crest, its wavelength, the distance between two wave crests, and its phase, the angle at a particular point along the wave's trajectory. The phases at the two round points at the top of each crest are the same.

When X-rays hit clouds of electrons orbiting atoms, they cause oscillation, which then leads to the emission of secondary X-rays in all directions. This is the process termed diffraction. When 'diffracted' X-rays are recorded by a detector, it is referred to as a reflection, because the angular relationship between the primary beam and the secondary beam is analogous to reflection of light by a mirror. As crystals consist of a regular 3D arrangement of the protein of interest, the diffraction pattern produced on the detector corresponds to the structure of the protein.

Rotation of the crystal exposes different reflecting surfaces (or 'planes', discussed below), producing different diffraction patterns, depending on the crystal orientation. Altering the crystal orientation in this way allows one to record a reciprocal 3D lattice of reflections, from which one can calculate the structure of the diffracting substance by use of the Fourier transform.

$$f_{hkl} = f_j e^{2\pi i(hx_j + kx_j + lz_j)} \quad \textbf{1}$$

The atomic structure factor, f_{hkl}, defines the contribution of a single atom, *j*, to a reflection with the indices of *h, k* and *l* in the reciprocal lattice, at positions x_j, y_j and z_j in the real lattice (see discussion of the 'unit cell' below). The scattering factor, f_j, is a mathematical representation of an atom's electron density which depends on the particular element. The integral sum of all atomic structure factors for the unit cell volume, V, results in the structure factor F_{hkl}:

$$F_{hkl} = \int_V \rho(x, y, z)\, e^{2\pi i(hx+ky+lz)}.\, dV \quad \textbf{2}$$

This equation can be rearranged to find ρ, the electron density:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)} \quad \textbf{3}$$

Relating this back to the properties of a wave: the amplitude and frequency of F_{hkl} are known as the amplitude is related to the reflection intensity and the frequency of a diffracted beam is usually equal to the incident beam. Therefore, the main hurdle to overcome in structure solution is the determination of the phase of F_{hkl} (160,161). The phase is the angle describing a particular point along a wave's trajectory. This problem is commonly referred to as the phase problem, and methods used to overcome this will be discussed below.

As the reciprocal and real lattices consist of a regular arrangement of points or reflections, then planes can be drawn across different points of the lattice, so that X-rays can be considered to be diffracted from planes within the crystal. A condition that must be satisfied to produce measurable detection is described by Bragg's law. Bragg's law states that a strong reflection will only be produced from these parallel planes, separated by interplanar distance d_{hkl}, if an incident beam strikes the planes at angle θ and is diffracted at the same angle, whilst satisfying the condition in equation 4:

$$2d_{hkl}.sin\theta = n.\lambda \quad \textbf{4}$$

Where θ is the angle of incidence, *n* is an integer and λ is the wavelength.

This can be explained further by observing Figure 2.1.6. Two X-rays entering the crystal are in phase with each other. But, due to the fact that the second beam travels further before being reflected at the same angle of incidence, the two X-rays leave the crystal out of phase. However, looking at Figure 2.1.6, if the distance between A to C to B, the geometrical path difference, is equal to an integral multiple of the wavelength, thereby equalling $2.d_{hkl}.\sin\theta$, then the two waves will be in phase and overlap to combine by constructive interference. Therefore, the addition of several waves in this manner produces a reflection strong enough to be recorded by the detector. Thus reflections in an individual diffraction pattern result from diffraction of many waves from parallel planes.

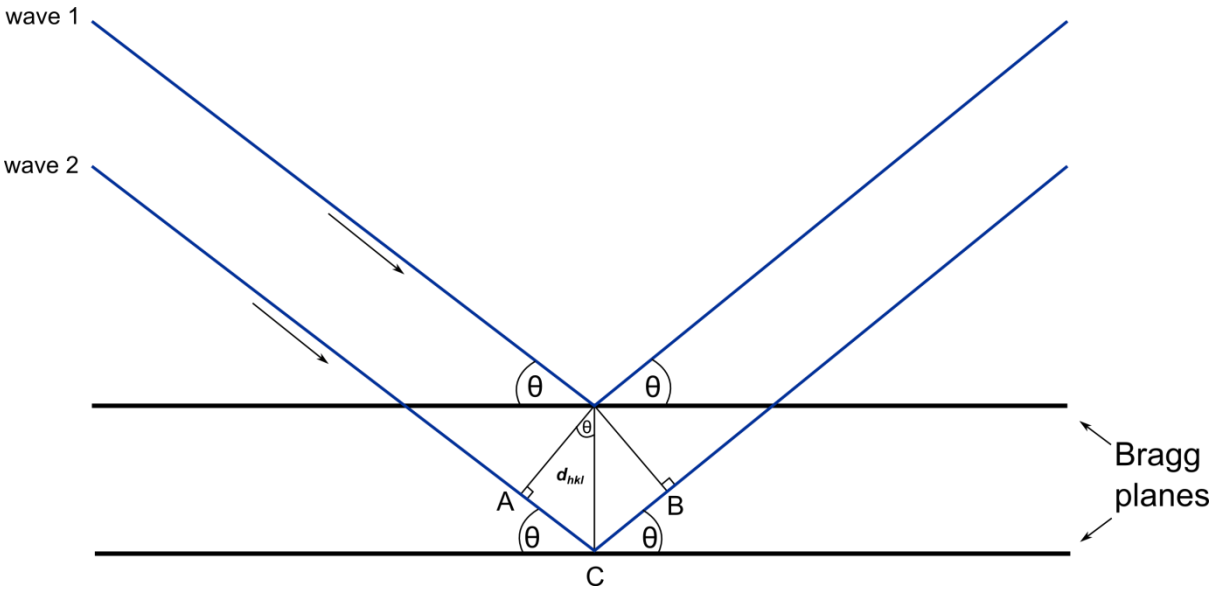


Figure 2.1.6 - Bragg's law illustrated by the diffraction of two parallel waves at angle, θ, from two parallel planes separated by d_{hkl}. If two waves enter the crystal in phase, they will leave in phase if the geometric path distance (i.e. the distance travelled by wave 2 before it becomes parallel with wave 1 again), ABC, is equal to an integer multiple of the wavelength, λ. Bragg's law is satisfied when the distance ABC = $n\lambda = 2.d_{hkl}.sin\theta$.

The angles of incidence (θ) differ based on the distance between sets of planes; with smaller distances (which produce higher resolution data) requiring larger angles. The intensity of each reflection depends on electron density, and thus, the identity and number of the atoms along a reflecting plane. An individual atom's contribution also depends on its fractional distance from the plane, as this affects the phase.

In summary, reflections are the sum of atom contributions along a plane, and include amplitudes and phases. Different reflections produced by the same atoms will have different phases, whilst amplitudes remain the same. Similarly, Friedel's law states that "the magnitudes of the structure factors of centrosymmetrically related reciprocal lattice points are equal"(160). Thus:

$$|F_{hkl}| = |F_{\overline{hkl}}| \quad \textbf{5}$$

These laws arising from symmetry are exploited later in solving crystal structures.

A deeper discussion of X-ray diffraction theory is beyond the scope of this thesis, but what has been described here should help to provide a context to the results presented herein.

TDH diffraction data were collected from synchrotron sources, the European Synchrotron Radiation Facility (ESRF) and Diamond Light Source (DLS). During data collection, crystals were mounted on a goniometer and maintained at low temperatures under a stream of nitrogen gas.

During data collection it is important to try and collect a complete data set. This requires that the crystal is rotated and that data is collected over a sufficient range to obtain an image of the full reciprocal lattice. The quality of data collected can be improved by collecting data over a larger area of the crystal, so that errors can be corrected for by having redundant data. Due to damage to the crystal caused by radiation over the course of the experiment, the extra data recorded may not be needed. In addition, if one fails to collect diffraction data when the crystal is in its necessary orientation, a full data set may not be achieved. For this reason, data collection can be performed by following a strategy, which stipulates the starting orientation of the crystal, the oscillation angle and the transmission of the incident X-ray required to make quality measurements on all unique reflections. Data collection experiments of TDH crystals were guided by the strategies given by the programs EDNA(162) or Mosflm(163), which determine the symmetry of the crystal to advise on the crystal orientation, oscillation angle, number of images and the X-ray transmission to avoid detector overloads. Alternatively, data were sometimes collected at small oscillation angles (0.5 – 1^0 per image) over 180-360^o to increase redundancy of the data.

2.1.3.2 Structure Solution and Refinement

As stated earlier, a crystal is a regular array of repeating subunits. In crystallography, these subunits are defined by the unit cell. The unit cell is defined by the axes a, b and c and the angles α, β and γ, as illustrated in Figure 2.1.7.

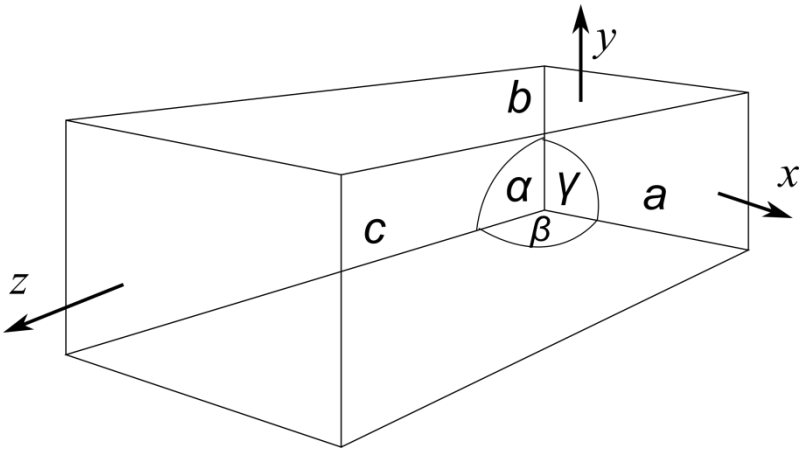


Figure 2.1.7 - a crystal unit cell. The unit cell is defined by the axes *a, b* and *c,* along the trajectories *x, y* and *z,* respectively, and by the angles α, β and γ.

Within the unit cell, there are often several copies of a protein and other molecules. The largest non-symmetrical unit that can be superimposed on other units in the cell is known as the asymmetric unit. The symmetrical arrangement of repeats of the asymmetric unit in the unit cell is described by the space group, which describes the combinations of rotational and translational symmetry that relate all the copies of the asymmetric unit to each other. A space group is denoted by a letter defining the lattice type and numbers that represent rotational and translational symmetry. For example, the common space group, $P2_12_12$, describes a unit cell, where the centres of symmetry are the vertices of the cell (denoted by P, a primitive lattice), and there are two screw axes (denoted by 2_12_1), which are axes which rotate the asymmetric unit by 180^0 and translate it by one half of the axis length, and one two-fold rotational axis (denoted by 2).

Early data processing was either carried out manually or by use of the integrated data reduction pipeline, xia2(164), at DLS. The first step in processing a crystallographic data set is integration of all the diffraction images. When early data processing was carried out manually, this was carried out using the graphical user interface (GUI) version of Mosflm, iMosflm(165). When automated processing was carried out by xia2, the program XDS was used. The process referred to as integration here actually requires three steps: autoindexing, postrefinement and integration. In iMosflm auto-indexing determines the lattice or unit cell parameters, crystal orientation and mosaic spread of the diffraction pattern using one image or a few images separated by several degrees of oscillation. Once the lattice type and unit cell parameters have been determined, spot (reflection) positions are predicted and are displayed to the user. The user can then choose the correct lattice type and unit cell parameters from a list based on probability scores and by visually checking the ability of each spot prediction to accurately locate spot positions. Postrefinement uses two separated ranges (wedges) of images and observes the distribution of spot intensity over partial reflections to refine unit cell parameters and improve the accuracy of the prediction. During integration the software takes the refined

unit cell and predictions of spot positions and counts the photons recorded at the predicted positions in each diffraction image, giving estimates of intensities.

Indexing is able to predict the unit cell geometry and lattice type, but not the space group, which describes the arrangement of the asymmetric units in the unit cell. The program POINTLESS(166,167) predicts the space-group of the crystal using a more thorough process than the integrating software. Reflections related to possible symmetry elements (e.g. a 2-fold rotation) are scored according to the probability that they are correct. Then, combinations of elements are combined and scored to detect the lattice symmetry. Space group symmetry is then determined by use of systematic absences. Systematic absences occur along screw axes, where symmetry related reflections cancel each other out. For example, in a 2_1 screw axis, there is an absence of intensity at every second measurement along the affected axis. Observing systematic absences was performed using the program POINTLESS, or was carried out manually to determine the space group from a number of options that share the same lattice symmetry.

During data collection, differences may arise between measurements of the same reflections. This is an issue, because it results in these measurements being on different scales. These differences or errors can come from many sources:

- The incident X-ray beam – e.g. variation in intensity, absorption by the crystal, variations in rotation speed and problems with shutter synchronisation;
- The secondary (diffracted) X-ray beam – e.g. absorption by the crystal and radiation damage to the crystal (resulting in weaker reflections);
- The Detector: detector calibration, the beamstop and beamstop support shadows.

In order to merge the data describing the same reflections together, the entire dataset needs to be scaled, which means to put the measured intensities on the same scale. Most of the factors described above are constant for any given image or at any given time in the experiment. Scaling applies correction factors (B factor and scaling factor) for each image or phi angle (the oscillation angle or orientation of the crystal). The knowledge of symmetry related reflections and Friedel's law is exploited here to determine how each intensity measurement should be corrected/scaled.

The programs Scala and its successor Aimless combine the processes of space group determination, scaling, merging, and conversion of intensities, I, to structure factor amplitudes, F. They were used during manual data processing and as part of the pipeline employed in automatic data processing by xia2. In all cases, the programs were set to exclude 5% of the data from processing to aid later analysis (see discussion of R_{free} below). The log files produced by these programs are very informative and report a number of data quality indicators, which will be outlined here because they were used to guide further refinement of the data.

R-factors give measures of differences between multiple observations of the same reflection and indicate the success of the scaling and merging process. They can also be used to decide

a resolution cut-off for data refinement and to compare scaling results between different space groups.

$$R_{merge} = \frac{\sum_{hkl} \sum_j \left| I_{hkl,j} - \langle I_{hkl} \rangle \right|}{\sum_{hkl} \sum_j I_{hkl,j}} \quad \textbf{6}$$

An ideal value for the R_{merge} is less than 10%, although data with an R_{merge} of up to 20% can be useful. The usefulness of this parameter is limited by the fact that it increases with multiplicity (redundancy of the data), which actually improves data quality. The multiplicity-weighted R-factor, R_{meas} improves on this – it reflects the higher accuracy achieved by having a greater number of measurements.

$$R_{meas} = \frac{\sum_{hkl} \left(\frac{N_{hkl}}{N_{hkl} - 1} \right)^{\frac{1}{2}} \sum_j \left| I_{hkl,j} - \langle I_{hkl} \rangle \right|}{\sum_{hkl} \sum_j I_{hkl,j}} \quad \textbf{7}$$

A newer parameter, the precision-indicating R-factor or $R_{p.i.m}$, decreases with increasing multiplicity and gives a better estimate of the precision of the scaled and merged intensity.

$$R_{p.i.m.} = \frac{\sum_{hkl} \sqrt{\frac{1}{N_{hkl} - 1}} \sum_j \left| I_{hkl,j} - \langle I_{hkl} \rangle \right|}{\sum_{hkl} \sum_j I_{hkl,j}} \quad \textbf{8}$$

The $R_{p.i.m.}$ is usually lower than the R_{merge} so a value of less than 10% was considered desirable when making decisions on data quality. Scala and aimless also present these parameters as calculated by batch of data, by intensity and by resolution.

The metrics (I)/sigma(I), and the mean (I)/sigma(I) (Mn((I)/sigma(I)) give an indication of the ratio between intensity, I, and the standard deviation, sigma, between related measurements. A high sigma, and thus a low (I)/sigma(I) may reflect error and such outliers may be rejected. A minimum Mn((I)/sigma(I)) of 2 for the overall data set was considered as an appropriate value for the data described here. As a lower (I)/sigma(I) may actually be valid for some reflections, especially high resolution reflections that refract weakly, a minimum (I)/sigma(I) of 1.0 was considered appropriate when choosing a resolution cut-off for data with an appropriate overall (I)/sigma(I). These metrics can also be useful in identifying areas of poor data quality, such as badly integrated data, poor images and radiation damage.

The completeness of a dataset is the proportion of all possible unique reflections that are recorded within the resolution limit. During data collection, data can be lost or missed due to radiation damage, a poor collection strategy or anisotropy. Completeness should ideally be as high as possible in all included resolution shells. A cut-off of 90% completeness in the outer shell (highest resolution data) was chosen to decide whether it was appropriate to retain data recorded to that resolution. For the datasets described here, most overall completeness values were close to 100%, as data was collected with the aim of achieving redundancy. In addition, data was collected in a manner to ensure that the multiplicity of the entire dataset was at least

2, as this is a value that one would expect for a complete data set, according to Friedel's law (see equation 5).

A newer measure, the half-dataset correlation coefficient was introduced in the program Aimless, the successor of Scala. This randomly splits each set of observations in two and compares the intensities by calculating a correlation coefficient (1.0 corresponds to a perfect correlation). This metric is independent of (I)/sigma(I), which is resolution dependent, so it is useful for picking a resolution cut-off.

After scaling and merging was complete, the program ctruncate was executed (automatically by aimless/scala) to convert the intensities into structure factors. Ctruncate estimates $|F|$ from I and sigma, using the average intensity in the same resolution range(168). The scaled and merged data are output in a .mtz file.

Once a scaled and merged dataset with structure factor amplitudes had been obtained, it was possible to begin building a model for the crystallised protein. Before a model can be created, it is useful to know how many protein subunits (e.g. monomers) are in the asymmetric unit. Most protein crystals have a solvent content in the range of 45-60%, with the solvent content of some proteins lying beyond this range – between 40-70%(169,170). Knowledge of the partial specific volume of a protein (the ratio between the volume of a protein and its molecular weight), the molecular mass of the protein subunit and the space group allow the iterative determination of solvent content and the number of protein subunits in the asymmetric unit. The CCP4 program Matthews_coef(171), will estimate the solvent content and the Matthews coefficient (V_M, ratio of crystal volume to protein mass), which should lie between 2.0 and 3.0 $Å^3.Da^{-1}$, using different values for the number of subunits in the asymmetric unit. The program outputs probability scores against each possible number of asymmetric units. The probability scores are calculated on the basis of a dataset of validated protein structures(169–171). The program Matthews_coef was used in this way to identify the likely number of TDH monomers in a unit cell. This number would be useful in the next stage of crystallographic model solution.

Despite having structure factor amplitudes at this stage, it is not possible to determine the absolute phases of the structure factors directly from an X-ray diffraction experiment. Phase solution, which enables the solution of a protein structure model, was carried out by molecular replacement (MR). The basic premise of MR is that a series of translation and rotation functions are applied to a 'search model', which is a 3D structure model of a similar or identical protein. MR was performed using the CCP4 programs, molrep(172,173) or Phaser(174).

Molrep performs molecular replacement by use of the Patterson function(175,176). To perform MR in this study, molrep was executed with the option to perform MR by using rotation and

translation functions. The .mtz file produced after earlier processing was used as input, and a .pdb file representing a previously solved *Tb*TDH structure was used as the "search model". The only additional non-default search parameter changed was the number of monomers in the asymmetric unit. This option tells molrep how many copies of the search model to search for in the asymmetric unit. Therefore, the value of this parameter was determined by the number of molecules in the asymmetric unit, as calculated using the Matthews coefficient (see above). For example, for a structure predicted to contain six TDH monomers in the asymmetric unit, a .pdb file containing one TDH monomer would be used as the search model and the program would be instructed to search for six copies of this. Alternatively, a search model that matched the predicted number of monomers in the asymmetric unit could be used. For example, dimeric TDH structures were used to perform MR on data where the structure was predicted to contain two molecules in the asymmetric unit.

The CCP4 program Phaser performs MR by a maximum likelihood method(177). Like molrep, phaser uses the scaled and merged .mtz file and a search model .pdb file as inputs. The particular .pdb files used as search models were chosen based on the same criteria as molrep. Phaser allows the user to define the percentage shared sequence identity between the search model and the target structure, so this value was set to 100. Two additional important parameters were used to perform MR with Phaser. The options under 'Define composition of the asymmetric unit' were used to instruct the protein of how many molecules of a certain molecular weight to search for. Alternatively, Phaser was instructed to perform the search looking for a specified number of copies of the search model in the asymmetric unit. This second option was particularly useful in some challenging cases. With structure models containing six monomers in the asymmetric unit, sometimes molrep failed to provide a full solution, and only provided solutions for 2 or 4 of the monomers present. Phaser was then useful in these cases, because two search models (called "ensembles" in Phaser) can be used to carry out MR simultaneously. An example of this involved instructing the program to search for one copy of a partial MR solution consisting of four monomers, whilst simultaneously searching for two copies of a .pdb file containing a TDH monomer. To develop initial models for TDH datasets, molrep was used in the majority of cases. Phaser was used in cases where molrep had failed to find a solution, or when it had only found a partial solution. After completing MR, the quality of the solutions were judged following refinement, as detailed below.

At this stage in structure solution, phases have been obtained, and the molecular positions of a model structure have been established. However, further refinement of atom positions is required. For this, an improvement in the accuracy of the calculated electron density map, which allows the discernment of fine molecular details, is desired.

Real-space refinement involves the iterative calculation of electron density maps from models. These electron density maps are then compared to the observed maps obtained from the data.

Refmac5, another program included in the CCP4 software package, performs reciprocal-space refinement: it aims to reduce differences between all parameters represented in the Fourier transform. Refmac5 uses a Maximum Likelihood method that incorporates Bayes' theorem, where likelihoods or probabilities are dependent on prior probabilities. In this context, the likelihood that a model is a correct representation of the structure that produced the data, is dependent on a number of parameters related to that particular protein, or to proteins in general. One example of this is the use of restraints. This increases the data-to-parameter ratio by restraining parameters to certain ranges of values (e.g. knowledge of amino acid geometry can be exploited to restrain main and side-chain torsion angles).

Refmac was run using the processed .mtz file and the .pdb file output by molrep or Phaser as inputs. As the TDH models often contained a non-protein molecule, NAD, a .cif file, which contains the restraints for that particular molecule, was also input into Refmac (see below for description of how .cif files were created for other ligands). Refmac was typically executed with the setting to run 10 cycles of restrained refinement of TDH models, producing a refined model and electron density maps. For later rounds of refinement, fewer cycles (5 or 3) were chosen, as additional cycles were observed to have little or no impact on indicators of the success of the refinement. Other default refinement parameters which told Refmac to refine isotropic B factors (see below for discussion of B factors) and to use automatic weighting were used. The default option to 'use experimental sigmas to weight X-ray terms' was turned off.

The success of refinement can be indicated by the calculation of the residual index, or R-factor:

$$R = \frac{\sum|(|F_o| - |F_c|)|}{\sum|F_c|} \qquad \mathbf{9}$$

where R is the residual index, and F_o and F_c are the structure factors from the observed and calculated data, respectively. The R-factor can be considered to be a measurement of the discrepancies between the observed and calculated data, and a lower number represents a lower discrepancy. A value of 50% or less suggests at least some similarity between the data. As the F_o and F_c data are used themselves to minimise the discrepancies between them, the use of the above equation to calculate R means that R is a biased measurement. For this reason, the same equation is applied to the 5% of data that was excluded from the refinement process. Therefore, Refmac reports R for the 'working' data set, R_{work}, and R for the 'free' dataset that was excluded from use in refinement, R_{free}. A well-refined model of data at resolution around 2.0Å, should give an Rwork of around 20%. The R_{free} is usually 5-10% above this. During refinement of TDH structures, an initial R_{work} of less than 30% for a MR solution was considered successful. For models in later stages of refinement, R_{work} values of around 20% were desirable. A large difference between the R_{work} and R_{free} can indicate an incorrect MR solution, large un-modelled areas of a structure, or over-fitting of the data, where areas of electron density caused by errors have been modelled with atoms. Therefore, a limit of 10% was chosen for the difference between R_{work} and R_{free} in a refined structure.

Refmac outputs a .mtz, from which two electron density maps can be calculated that allow the manual and visual manipulation of structure models. The two maps of use in achieving this are calculated:

- $2F_o$ minus F_c maps – by doubling F_o and subtracting F_c, areas where atoms have been placed correctly will have less electron density, and areas where atoms should be located should have greater density.
- $F_o – F_c$ maps ('difference maps') – positive electron density will show in areas where atoms are missing, whereas negative density will appear where atoms have been placed in error. A perfect structure solution would produce an empty difference map.

These maps can be displayed according to the root mean standard deviation (RMSD or σ), which is related to the probability of attaining a certain structure factor amplitude, given the model. By setting the contours of electron density maps to a certain σ level, one can choose to place atoms with a specified degree of confidence.

The structure visualisation and manipulation software, Coot(178,179), was used on Windows and Linux systems to carry out real space refinement and validation of structures. Manipulation of structures was guided by $2F_o$-F_c maps at σ levels of 1.5 and by F_o-F_c maps at σ levels of 2.00. The aim here is to achieve the best possible fit of the model to the electron density map. A number of tools within the Coot program were used during the refinement process. Firstly, the 'Real space refine zone' was used to automatically manipulate a defined region of the protein to improve the fit to the electron density map. Alternatively, the 'Rotate Translate Zone/Chain/Molecule' tool was used to manually manipulate the positions of atoms or defined regions of the protein to improve the fit in a more controllable way. Manipulating protein structures in this way creates the risk that the model may take on biologically implausible orientations. To counteract this, the tool 'Regulate Zone' automatically changed the orientation of a defined zone of the protein so that it conformed to the Refmac restraints defined earlier. Other tools used to quickly fix this problem include the 'Flip Peptide' tool and the 'Change reidue's phi and psi angles' tool, which were used to alter the geometry around peptide bonds. Some amino acids, such as asparagine, glutamine and histidine are easily placed in the wrong orientation, due to their planar nature. Using the apparent hydrogen-bonding networks as a guide, the "Auto-flip sidechain' tool was used to flip the residue's side chains by 180° when the orientation appeared to be incorrect. On some occasions, determining the orientation of more disordered parts of the molecule, where density was sparse, was not possible using the aforementioned tools alone. Therefore, the residues in these regions of the model were deleted and re-added using difference maps as a guide, following multiple successive rounds of refinement of the model using Refmac. In this case, the residues were added as L-alanine and then converted to the correct residue using the 'mutate' or 'mutate & auto-fit' tools.

Two important parameters that are included in the PDB files are the B-factor (or temperature factor) and occupancy. The B-factor is a measure of disorder in a crystal structure, with higher numbers indicating a greater degree of variation in the coordinates of an atom. The occupancy

of an atom describes the degree to which a certain space in the unit cell is occupied by an atom, with full occupancy being represented by 1.0. Altering the occupancy and B-factor can have a similar effect on protein models. Altering B-factors is more appropriate for most macromolecular structure refinements, and thus was used to account for the variation of atom positions within adjacent areas. Occupancy was used to model atoms which may not always be present in every asymmetric unit, for instance ligands which may be present in only a proportion of asymmetric units. It is also possible to model atoms in 'multiple occupancies', and this was carried out when it was evident from the electron density map, that an atom or group of atoms occupied two or more distinct positions.

To incorporate ligands into the model, structures were produced as .pdb files, along with restraints (stored in .cif files, used by Refmac), using the online ProDRG server(180). Once placed in the desired area, the ligand was merged with the structure so that it was stored in the same PDB file. It is common for solvent, precipitant or cryoprotective molecules to be found in structures. Coot has a feature that was frequently used in this study to model several of these molecules into structures. Restraints for these molecules are contained in the default Refmac restraint dictionary. Water molecules were modelled into roughly spherical regions of electron density as oxygen atoms. This was done automatically using Coot's 'find waters' tool and manually by placing atoms.

After each round of manipulation, the modified model, the .mtz file containing the structure factor amplitudes, and any .cif restraint files for ligands were input into Refmac5 before executing, to produce a new pair of electron density maps, along with a new, modified model.

Towards the end of structure solution, it is vital to validate the model, using knowledge of protein structure, chemical bond geometry and ligand structures. Bond distances and geometries, including those in hydrogen bonding networks, were observed visually and modified accordingly. Coot has a number of useful tools for structure validation, such as the Ramachandran plot. The Ramachandran plot can be used to identify amino acids with phi (φ) and psi (ψ) angles (see Figure 2.1.8), that lie outside of normal values. This plot was used in this way to find and correct such errors.

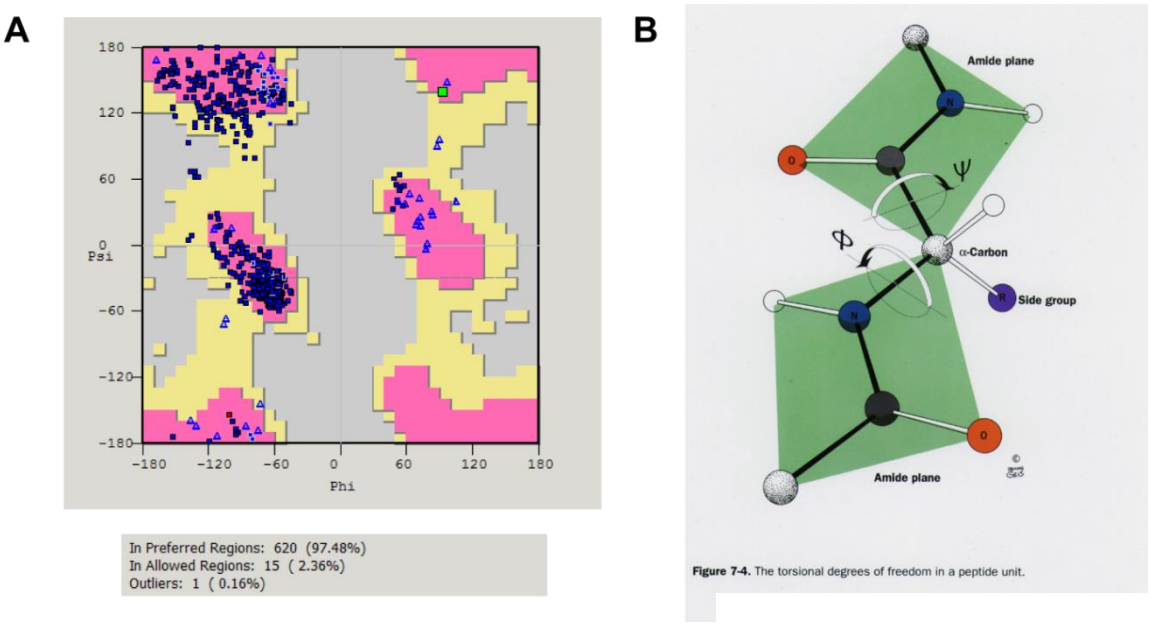


Figure 2.1.8 - the Ramachandran plot (A) as a structure validation tool. The planarity of peptide bonds, and the presence of amino acid side chains restrain the angles on either side of the amino acid α-carbon. Thus phi (ϕ) and psi (ψ) are constrained to certain combinations of values (see image B). These values appear on Ramachandran plots (see image A) as 'preferred regions' and 'allowed regions'. Residues with values outside of these regions of the plot ('outliers') suggest an incorrect geometry. Note: the lack of a side chain on glycine residues mean that they often lie outside of the preferred and allowed regions for other amino acids.

Another tool within Coot that was exploited to help validate and correct models was the rotamer analysis tool. This tool observes the combination of torsion angles between amino acid side chains and highlights the residues where a combination has a low probability of existing in nature. The rotamer analysis tool was used to correct side chain orientations in TDH structures.

2.1.3.3 Sequence and Structure Analysis

The Basic Local Alignment Search Tool (BLAST) is a suite of programs that are executed to generate alignments between query amino acid or nucleotide sequences and other sequences in the database. This can provide an insight into the similarity between the query protein and proteins from other organisms. The amino acid sequence and nucleotide sequence (both provided as single letter representations), for TDH and the TDH gene from *T. brucei gambiense* (Tb927.6.2560), respectively, were obtained from the TriTryp database(155). Searches against these sequences were executed using the National Center for Biotechnology Information (NCBI) protein and nucleotide BLAST tools(181). Sequence alignments were reported along with sequence identity matches (percentage), and the organism name. Results of biological or pharmacological interest were noted.

Protein structure models and their interactions with ligands were observed visually using PyMOL (Schrödinger LLC), Coot and UCSF Chimera(182). To analyse the average B-factor for main chain atoms of TDH amino acid residues, nine representative monomer structures were

taken from different PDB files. The CCP4 program baverage was then used to calculate the mean B-factor value across all of these structures. To visualize the B-factor variance, the values output by baverage were input into an attribute file, which is used by UCSF Chimera to assign new attributes to a protein structure. The attribute file, containing the average B-factors for each amino acid residue, was assigned to a structure of a TDH monomer. The 'colour by attribute' feature of UCSF Chimera was then used to colour a ribbon representation of TDH by the average B-factor. To measure the disorder of TDH by a different method, the same nine representative structures were superimposed using the 'Matchmaker' and 'match→align' tools in UCSF chimera. The RMSD for the α-carbons (Cα) of each residue was then computed. The RMSD values were output and used in the same way as the average B-factors to visualize the variation in positions of amino acid residues.

The program PISA (Protein Interfaces, Surfaces and Assemblies)(183) in the CCP4 package was used to output the quaternary structures of TDH models. This is done using the atomic coordinates from the PDB file and the space group, to generate all symmetry-related protein models.

2.1.4 Other Biochemical and Computational Techniques

2.1.4.1 Size-exclusion Chromatography

Size-exclusion chromatography or gel filtration is a method of distinguishing between proteins of different molecular mass, by virtue of their differential migration speeds through a polymer gel. The schematic in Figure 2.1.9 outlines how migration through pores of certain size and tortuosity will determine the migration speeds of proteins through a gel-packed column. Prior to an experiment, a gel filtration column must be calibrated with proteins of known molecular mass. Then a linear gradient can be calculated for the relationship between their molecular masses and the volume of buffer required to elute them from the column (elution volume). A protein or mixture of proteins may then be applied to such a gel. Then, the elution volumes of the proteins can be used to estimate the molecular weights of the eluted molecules. This can potentially inform on the quaternary structures of proteins and the existence of multi-protein complexes.

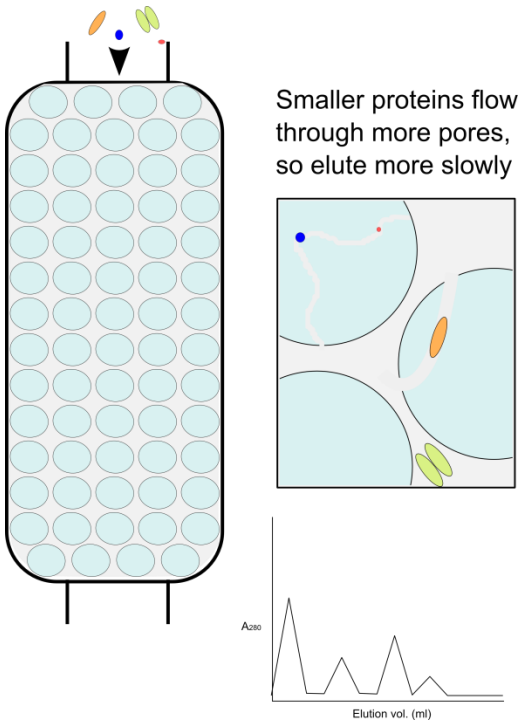


Figure 2.1.9 - Size-exclusion chromatography schematic illustrating the migration of proteins through a gel packed column (left and top right). The elution of proteins as measured by UV absorbance at 280nm is used to estimate molecular mass.

The sample buffer used for gel filtration was similar to the storage buffers used for TDH and KBL, with an intermediate NaCl concentration: 50mM Tris.HCl, 200mM NaCl, 1mM βM, 10% v/v glycerol, pH 7.5. This buffer was cleaned and de-gassed by Buchner filtration. First, the pumps of the FPLC machine (Controller LCC-501 plus) were lubricated with 20% ethanol and then washed with sample buffer. A Superdex 200 10/300 GL column (Amersham Pharmacia Biotech, part of GE healthcare, Buckinghamshire, UK) was equilibrated with approximately 2 column volumes of sample buffer. All samples were centrifuged for 5 minutes to remove particulate matter prior to loading. 400µl samples were loaded with a syringe and the eluent was collected in 2ml fractions. Samples were run at a flow rate of 0.5ml/min and a maximum pressure of 1.5MPa. Before any experimental samples were run, the column was calibrated with the standards: β-amylase (MW 200 kDa), alcohol dehydrogenase (MW 150 kDa) bovine serum albumin (BSA) (dimer MW 132 kDa; monomer MW 66 kDa), ovalbumin (MW 45 kDa), carbonic anhydrase (MW 24 kDa) and cytochrome C (12.4 kDa). A number of different experimental samples containing mixtures of different proteins and ligands were tested. The various samples can be summarised as follows:

- TDH alone;
- TDH with ligands;
- KBL alone;
- KBL with ligands;
- TDH and KBL;
- TDH and KBL with ligands.

Control samples containing ligand only were also tested in case any of them interfered with the detection method. The elution of protein was monitored by UV absorbance at 280nm by a UV M-II spectrophotometer (GE Healthcare). Occurrence of retention peaks was recorded by a REC 112 chart recorder (GE Healthcare). To calculate the void volume (V_0), the minimum volume required to elute a protein from the column, which corresponds to the volume of fluid between the polymeric beads of the gel, a solution of blue dextran (MW 2000kDa) was ran. Blue dextran is so large that it does not flow through the pores of the gel beads, and thus elutes at the V_0.

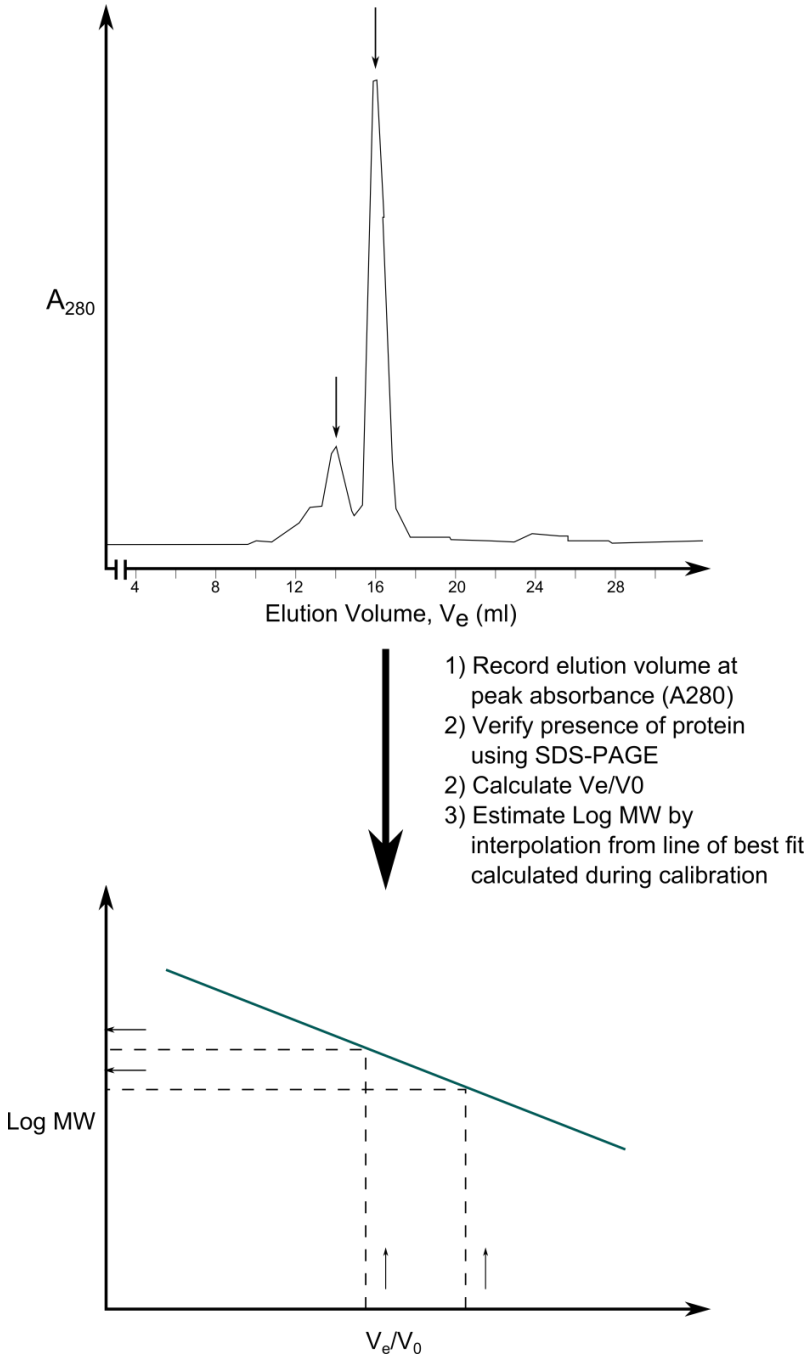


Figure 2.1.10 - schematic illustrating the method used for estimating molecular weight by size-exclusion chromatography.

To calculate the molecular weights of proteins eluting at a particular elution volume (V_e), a plot of $\log_{10}$ molecular weight of standard protein against elution volume divided by void volume (V_e/V_0) was made using Microsoft Excel. A line of best fit was calculated for the points and the gradient of the line was used to estimate the molecular weights of proteins eluting at certain volumes, using the formula, y = mx + c.

2.1.4.2 Cross-linking

Another method of indicating the quaternary structure of proteins and for the identification of multi-protein complex formation is the use of cross-linking reagents. These reagents react with amino acid side-chains on molecules that come into close contact with each other. This maintains and stabilises the interaction, so that the sizes of complexes can be monitored by SDS-PAGE, which would usually disrupt the interaction. Cross-linking studies of TDH and KBL were carried out using a method similar to that of Davies and Stark(188), which employed dimethyl suberimidate (DMS) as the cross-linking agent. DMS reacts with free amines on lysine residues to form covalent linkages. The reactions predominate within oligomers, although non-specific cross-linking can become apparent after long periods of time.

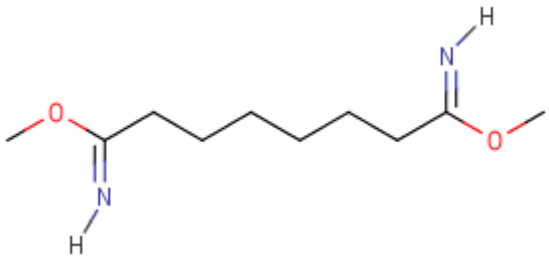


Figure 2.1.11 - the chemical structure of dimethyl suberimidate.

Protein concentrations and DMS concentrations were carefully chosen to avoid precipitation. A number of different experiments were carried out to test:

- The presence of a multi-enzyme complex between TDH and KBL
- The quaternary structures of TDH and KBL
- The effect of enzyme concentration on enzyme oligomerisation
- The effect of temperature on oligomerisation
- The effect of substrates on enzyme oligomerisation

To confirm the quaternary structures of TDH and KBL, to test the effect of enzyme concentration and to detect the existence of a multi-enzyme complex, the following conditions were tested:

- TDH (2.6×10^{-2}mM [1mg/ml]) + DMS (7.3 mM [2mg/ml])
- TDH (1.3×10^{-2}mM [0.5mg/ml]) + DMS (7.3 mM [2mg/ml])
- TDH (1.3×10^{-2}mM [1mg/ml]) control
- KBL (2.2×10^{-2}mM [1mg/ml]) + DMS (7.3 mM [2mg/ml])
- KBL (1.1×10^{-2}mM [0.5mg/ml]) + DMS (7.3 mM [2mg/ml])
- KBL (2.2×10^{-2}mM [1mg/ml]) control

- TDH (2.6 x 10^{-2}mM [1mg/ml]) + KBL (2.2 x 10^{-2}mM [1mg/ml]) + DMS (14.6 mM [4mg/ml])
- TDH (1.3 x 10^{-2}mM [0.5mg/ml]) + KBL (1.1 x 10^{-2}mM [0.5mg/ml]) + DMS (14.6 mM [4mg/ml])
- Bovine Serum Albumin (1.5 x 10^{-2}mM [1mg/ml]), DMS (7.3 mM [2mg/ml]) positive control

All solutions were made up to 100µl in 200mM Tris.HCl buffer at pH 8.5. Final protein concentrations were approximately 1mg/ml Reactions were carried out in duplicate, at room temperature and at 2-8°C. Reactions were incubated overnight and the reactions were halted by the addition of Laemmli buffer in preparation for analysis by SDS-PAGE. Gels of 9% polyacrylamide, rather than the usual 12%, were used to increase the migration of higher molecular weight complexes.

To test the effect of TDH substrates on its oligomerisation, the following conditions were tested:

- TDH (2.6 x 10^{-2} mM [1mg/ml]) + DMS (7.3 mM [2mg/ml])
- TDH (2.6 x 10^{-2} mM [1mg/ml]) + 1mM NAD^+ + DMS (7.3 mM [2mg/ml])
- TDH (2.6 x 10^{-2} mM [1mg/ml]) + 10mM NAD^+ + DMS (7.3 mM [2mg/ml])
- TDH (2.6 x 10^{-2} mM [1mg/ml]) + 1mM NAD^+ + 30mM L-threonine + DMS (7.3 mM [2mg/ml])
- TDH (2.6 x 10^{-2} mM [1mg/ml]) + 10mM NAD^+ + 30mM L-threonine + DMS (7.3 mM [2mg/ml])
- TDH (2.6 x 10^{-2} mM [1mg/ml]) + 10mM NAD^+ + 15mM L-threonine + DMS (7.3 mM [2mg/ml])
- 10mM NAD^+ + 30mM L-threonine + DMS (7.3 mM [2mg/ml]) control

These experiments were carried out as above, but with a time course limited to 3 hours.

To test the effect of KBL substrates on the oligomerisation of that enzyme, the following conditions were tested:

- KBL (2.2 x 10^{-2} mM [1mg/ml]) + DMS (7.3 mM [2mg/ml])
- KBL (2.2 x 10^{-2} mM [1mg/ml]) + 0.5mM pyridoxal phosphate (PLP) + DMS (7.3 mM [2mg/ml])
- KBL (2.2 x 10^{-2} mM[1mg/ml]) + 15mM glycine + DMS (7.3 mM [2mg/ml])
- KBL (2.2 x 10^{-2} mM [1mg/ml]) + 2.5mM acetyl-coenzyme A + DMS (7.3 mM [2mg/ml])
- KBL (2.2 x 10^{-2} mM [1mg/ml]) + 0.5mM PLP + 15mM glycine + DMS (7.3 mM [2mg/ml])
- KBL (2.2 x 10^{-2} mM [1mg/ml]) + 0.5mM PLP + 2.5mM acetyl-coenzyme A + DMS (7.3 mM [2mg/ml])
- KBL (2.2 x 10^{-2} mM [1mg/ml]) + 0.5mM PLP + 2.5mM acetyl-coenzyme A + 15mM glycine + DMS (7.3 mM [2mg/ml])
- KBL (2.2 x 10^{-2} mM [1mg/ml]) + 15mM glycine + 2.5mM acetyl-coenzyme A + DMS (7.3 mM [2mg/ml])
- 0.5mM PLP + 2.5mM acetyl-coenzyme A + 15mM glycine + DMS (7.3 mM [2mg/ml]) control

These experiments were carried out over 3 hours, as for the previous set, though reaction volumes were decreased to 50µl. The volume was reduced to preserve resources because the reaction volume had no demonstrable effect on results, and only a small sample of the reaction mixture (10µl) was required for SDS-PAGE analysis.

2.1.4.3 Pull-down Assay

The pull-down assay is a simple method of testing for binding interactions between two or more proteins. The pull down assay relies on the binding of a (usually, tagged) protein to a resin. A second, untagged protein is added to the resin. After washing the resin with a wash buffer to remove unbound proteins, an elution buffer is used to wash all bound proteins from the resin. If the second protein is recovered with the eluted first protein, this suggests that there is a binding interaction between the two proteins, as it would otherwise have been removed with wash buffer. The first step required was the removal of the poly-histidine tag from TDH. The pET-15b vector encodes an amino acid sequence between the His-tag and the protein that is recognised and cleaved by thrombin. 1ml of TDH solution was incubated overnight in solution with approximately 5 units of thrombin, to give a ratio of 1 unit of thrombin to 1mg of protein. This solution was then passed through a HisTrap® nickel column connected to a benzamidine column, to remove the cleaved histidine residues and the thrombin. Cleaved TDH passed through the column. For the pull-down assay, 1ml of KBL solution (13.1 X 10^{-2} mM [6mg/ml]) was loaded onto a nickel column. The cleaved TDH solution was then loaded onto the column at a flow rate of approximately 0.5ml/min. After incubating the column for 60 minutes, the column was washed with 10-20 volumes of wash buffer. Then 5 column volumes of elution buffer were added to elute the bound protein(s). The success of the procedure was determined by SDS-PAGE.

2.2 Enzymatic Studies

2.2.1 Basic Kinetics

An integral part of this research effort was the study of the enzymatic activity of the enzymes studied. The activity may be affected by numerous factors, such as the concentration of the enzyme's substrates and co-factors, and the pH of the environment. Detailed experiments are necessary to gather enough information to characterise these effects in a way that is useful for describing the behaviour of the enzyme. For this reason, it is equally important that a lot of attention to detail is applied to the design of assays.

Some of the principal considerations are the mode of detection of enzyme activity and the type of assay. For TDH assays, detection of enzyme activity is aided by the difference in absorbance of ultraviolet radiation (UV) at 340nm by the cofactor NAD^+ and its reduced form

NADH. Thus as the reaction proceeds, it can be monitored by the increase in absorbance at 340nm (see Figure 2.2.1)

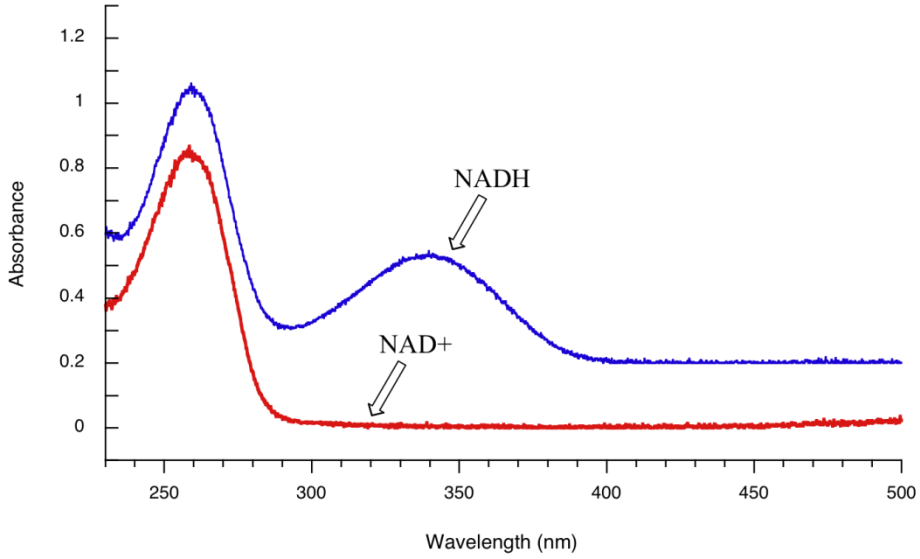


Figure 2.2.1 - The absorbance of NAD^+ and NADH at 340nm. The greater absorbance of by NADH at 340nm, relative to NAD^+, can be used to measure the reaction of TDH. Image obtained from http://chemwiki.ucdavis.edu/Organic_Chemistry/Organic_Chemistry_With_a_Biological_Emphasis/Chapter_04%3A_Structure_Determination_I/Section_4.3%3A_Ultraviolet_and_visible_spectroscopy.

All enzyme assays were carried out using an UltroSpec 3000 (GE Healthcare) UV/Vis spectrophotometer or a FLUOstar Omega plus plate reader (BMG Labtech). Enzyme assays are typically carried out by use of an end-point assay, where measurements are made at the beginning and end of an assay, or a kinetic assay, where measurements are made at multiple time-points to establish a rate of reaction. For the studies described here, time-point assays were chosen for a number of reasons. Firstly, time-point assays can be more simplistic because they do not require the addition of extra reagents to halt the reaction at a specific time-point. Secondly, time-point assays can help to reduce error – by calculating a gradient from measurements made over several time points, the effects of errors or artefacts (caused by bubbles, dirt, etc.) can be minimised. In a kinetic assay, it is important to measure the initial velocity of enzyme activity, where the steady state assumption holds true(189). At this time it is assumed that the reaction has reached equilibrium, but that the substrate concentration is so much higher than that of the enzyme and bound intermediates that the relative concentrations remain constant. The increase in product concentration should be linear with time when the steady state assumption is true. However, after some time, this linearity is lost as the substrates begin to become depleted. Other phenomena that can cause error are present nearer the initiation of a reaction. In some instances, there may be a lag before the enzyme reaches steady-state velocity. Conversely, the reaction may begin with very high activity and then decrease to a second linear phase of lower activity. The first phase is called the 'burst phase' and is believed to be caused by the fact that the enzyme may be bound to a substrate or

cofactor prior to reaction initiation, so there is a 'burst' in enzyme activity, as the binding of substrates has already occurred. These considerations are illustrated in Figure 2.2.2.

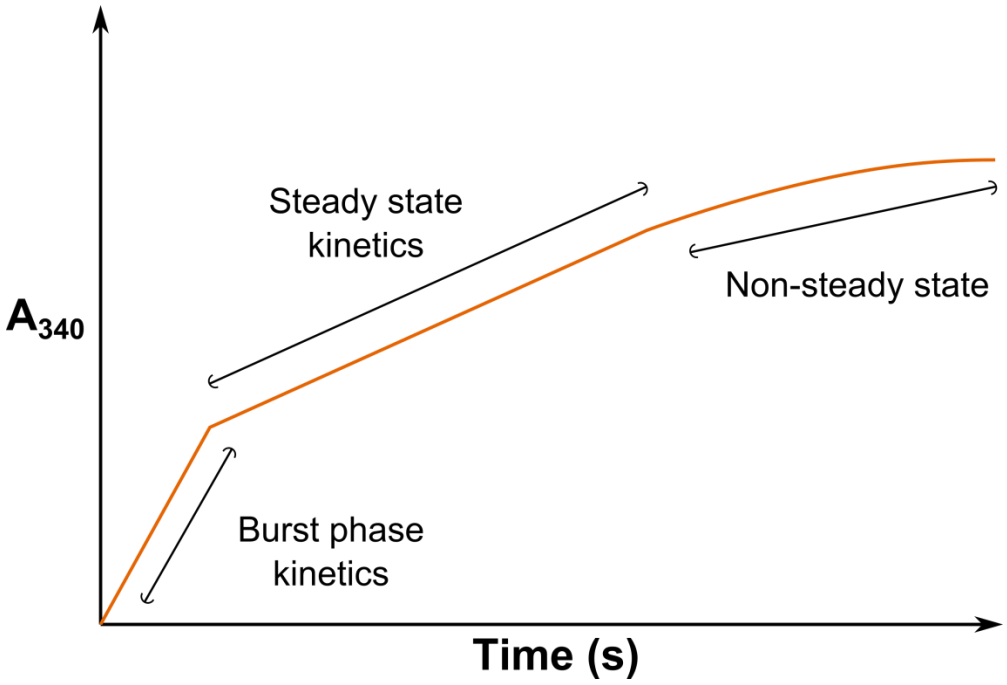


Figure 2.2.2 - an example of a progress curve during a kinetic enzyme assay. Measurement should be delayed for the first few seconds, to avoid the 'burst' of activity that does not represent steady state kinetics. Conversely, the earliest start of the progress curve can show a lag in activity. Measurements should also be avoided too late in the assay, where the depletion of substrates means that the steady-state assumption is no longer true. Measurements should be made at the linear part of the curve, which represents the enzyme at the steady state.

In a kinetic assay, it is also important that the mode of detection has a large enough signal window to measure differences in activity. However, it is also important that the signal produced is valid and does not saturate the detector. In spectrophotometric absorbance assays, light absorbance, which is a logarithm of light transmittance, is reported. An absorbance measurement of 1.0 is equal to 10% of light transmittance and a measurement of 2.0 is equal to 1% of light transmittance. Depending on the sensitivity of the detector, it is desirable for the final measurement to not be significantly above an absorbance of 1.0. The calculation of NADH concentration relies on the Beer-Lambert law(190):

$$A = \varepsilon bc \rightarrow c = \frac{A}{\varepsilon b} \quad \textbf{10}$$

using the absorbance, A, the pathlength, *b* (e.g. 1cm), and the extinction coefficient, ε, of NADH, 6220mol^{-1}. Deviations from this law can occur as the absorbance rises beyond 1.0. Therefore, for TDH assays carried out in the Ultrospec 3000, maximum absorbance measurements were aimed to be close to 1.0, whereas for assays carried out using the FLUOstar Omega plate reader, maximum measurements between an absorbance of 1.0 and 2.0 were deemed appropriate. This was achieved through choosing an appropriate enzyme concentration.

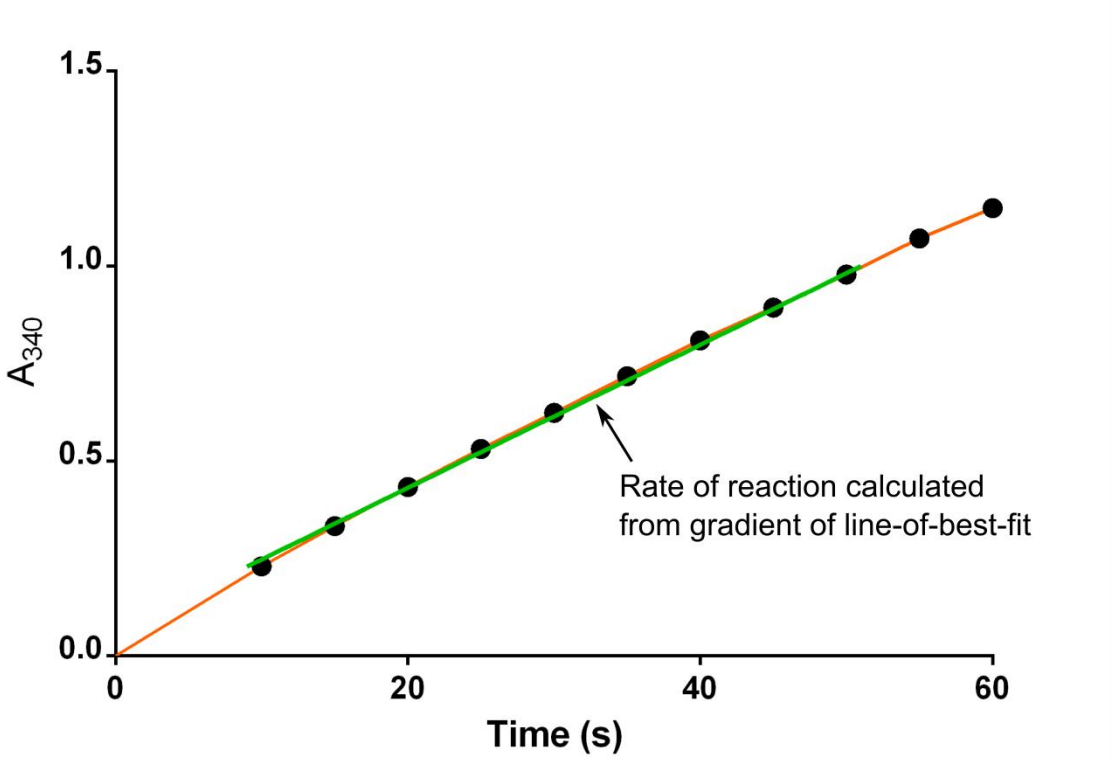


Figure 2.2.3 - an example of raw data collected during a typical kinetic assay. Each data point is collected by an orange line, and the green line is the line-of-best-fit, from which the gradient is used to calculate the velocity of catalysis.

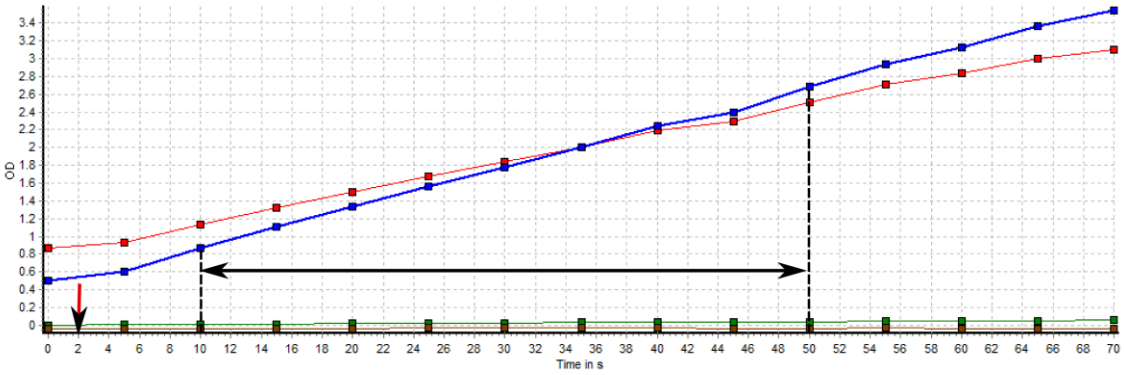


Figure 2.2.4 - an example of raw data collected during a typical kinetic assay using the FLUOstar OMEGA plate reader. The blue points represent absorbance (or optical density, OD) at 340nm, whilst the red points show absorbance at 320nm. The pale green and brown points do not show an increase in absorbance as they represent absorbance values at 400nm and 500nm, respectively. The rate of reaction was calculated from the linear gradient of the line fit to the relevant data points, as described in Figure 2.2.3 and Figure 2.2.5.

To help determine the composition of the reaction mixture, assay design was guided by the TDH assays of McGilvray & Morris(191) and Boylan & Dekker(120). The reaction buffer used contained 200mM Tris.HCl, adjusted to pH 8.5. Preliminary assays of TDH activity were carried out to identify saturating substrate concentrations, the mechanism of action of the enzyme, and statistics such as the half-saturating substrate concentration, K_M, and the maximal rate of reaction, V_{max}. Several reactions were performed over a range of fixed concentrations of one

substrate whilst varying the other. The range of NAD^+ concentrations tested was 0.1, 1, 5, 10, 15 and 20mM. The range of L-threonine concentrations tested was 1, 5, 10, 20, 30 and 40mM.

100mM stock solutions of L-threonine and NAD^+ were made by dissolving the appropriate masses in reaction buffer. It was observed that NAD^+ spontaneously degraded over prolonged periods of time. Such degradation, which could have affected the results, was detected by an increased absorbance by the stock solution at 340nm. This change was not considered appreciable within the first 24 hours, so NAD^+ stock solutions were always made on the same day of the experiment or on the previous evening. The reaction mixture was made by adding reaction buffer, L-threonine and NAD^+ up to a total volume of 980µl in a 1cm path length quartz cuvette. This mixture was incubated in a Grant SUB14 water bath (Grant Instruments) at 37°C (310K) for 5 minutes prior to initiating the reaction. The enzyme was diluted with reaction buffer down to a concentration of 0.66mg/ml (17.5 µM), which was checked using the NanoDrop ND-1000 (Nanodrop, part of Thermo Fisher Scientific, Leicestershire, UK), and it was this enzyme solution that was added to the reaction mixture (final concentration 0.0132mg/ml [3.5 µM]). Prior to initiation of the reaction, the opening of the cuvette was covered with Parafilm and the cuvette was inverted seven times to ensure thorough mixing. Then, 20µl of enzyme solution was added to the reaction mixture, with stirring, to initiate the reaction. The conversion of NAD^+ to NADH was monitored at 340nm in an Ultrospec 3000 UV/Vis spectrophotometer (GE Healthcare), taking measurements at 5 second intervals, beginning 10 seconds after the start of the reaction to allow for equilibration, up to 60 seconds. Only measurements up to 50 seconds were considered in analyses as the increase in absorbance was deemed to be consistently linear up to this time across all tested conditions. The catalytic rate was determined according to the Beer-Lambert law(190),using the gradient of a line of best fit from the measured absorbance, *A*. Control experiments were run as above, but omitting the enzyme, L-threonine or NAD^+. The spectrophotometer was always blanked on the full mixture on the basis of the observation that there was some background absorbance when the enzyme was present, but one of the substrates was omitted. It seemed possible that NAD^+ co-purified with TDH so small amounts of activity were observed when NAD^+ was excluded. However, this activity was judged to be too small to significantly affect the results, considering the design of the assay (see discussion of burst phase kinetics above).

Once saturating concentrations of NAD^+ and L-threonine concentration were established at 10mM and 30mM, respectively, more detailed studies were carried out at a saturating concentration of one substrate, whilst varying the concentration of the other substrate. TDH activity was assayed at 1, 2.5, 5, 10, 12.5, 15, 20, 22.5, 25, 27.5 and 30mM L-threonine, and at 0.1, 0.5, 1, 2, 3, 5, 7.5 and 10mM NAD^+. The assays were carried out as detailed above, at both 37°C and at room temperature (23°C).

All preliminary assays were repeated up to a maximum of five times (on separate days) until three measurements of rate, *v* within a standard deviation of 10% were obtained. If five values were obtained without this, the values that had the lowest standard deviation when averaged (at

least three measurements) were included in the analysis. Kinetic constants were derived by use of non-linear least squares fitting to the Hill modification of the Michaelis-Menten equation(192–194):

$$v = \frac{V_{max}[S]^h}{{K_M}^h + [S]^h} \quad \textbf{11}$$

where *v* is the rate of reaction, V_{max} is the maximum rate, [S] is the substrate concentration, K_M is the Michaelis constant (the substrate concentration that produces a half-maximal rate of reaction) and *h* is the Hill coefficient, which represents the minimum number (to the nearest integer) of active sites involved in cooperative interactions. A value of $h = 1$ suggest there is no cooperativity and a value of $h < 1$ or $h > 1$ represents negative or positive cooperativity, respectively. Positive cooperativity is usually manifest in sigmoidal *v* versus [S] curves. A test for cooperativity was applied using the following equation:

$$h = \frac{\log 81}{\log(\frac{[S]_{0.9}}{[S]_{0.1}})} \quad \textbf{12}$$

$[S]_{0.9}$ and $[S]_{0.1}$ are the substrate concentrations at 90% of the V_{max} *and 10% of the* V_{max}, respectively. It is a useful indicator of non-hyperbolic increases in *v* with [S](194). All analysis of this data was carried out using the following data analytical software: Microsoft Excel (2010), Origin Pro (Origin Lab, Massachusetts, USA) and Prism (version 6 for Windows, 2014, Graphpad Software, California, USA, www.graphpad.com).

Figure 2.2.5 summarises the data analysis process of enzyme kinetic data, as described above. The raw data consists of measurements of the absorbance over different time points over the course of the experiment. In Microsoft Excel, the rate of increase in absorbance was calculated from the gradient of a line-of-best-fit. This value was then converted to the rate/velocity of catalysis by use of the Beer-Lambert law (see Equation 10). The mean velocity values and standard deviations were then calculated from the results of repeat experiments in Microsoft Excel. To determine kinetic parameters (e.g. K_M, V_{max}), the mean values and standard deviations were input into either Origin Pro or Prism software, which plotted the values on a graph. These programs calculate the appropriate kinetic parameters by non-linear regression analysis by the least-squares method. For example, the programs were used to fit curves described by an equation, such as Equation 12, to the plotted data. The fit of the curve to the data is improved by iterative modification of the unknown parameters, until the squared differences between the curve and the data points are minimised. To illustrate this further, consider the use of the Michaelis-Menten equation(195,196) by Prism:

$$Y = Vmax*X/(Km + X)$$

The value X is the substrate concentration, [S], and Y is the velocity of catalysis, *v*. Both of these values are determined by the data, and so are fixed. The data analytic software input the values of X into the equation and automatically modify the values of unknown parameters until

the differences between experimentally-derived Y values and calculated Y values are minimised. The resultant plots consist of data points, error bars and the fitted curve. Note that the error bars on the plots presented in this thesis relate to standard deviation (SD) values, so even when the SD relative to *v* is lower at higher values of *v*, the absolute value of SD is plotted, so larger error bars are often seen at higher values of *v*.

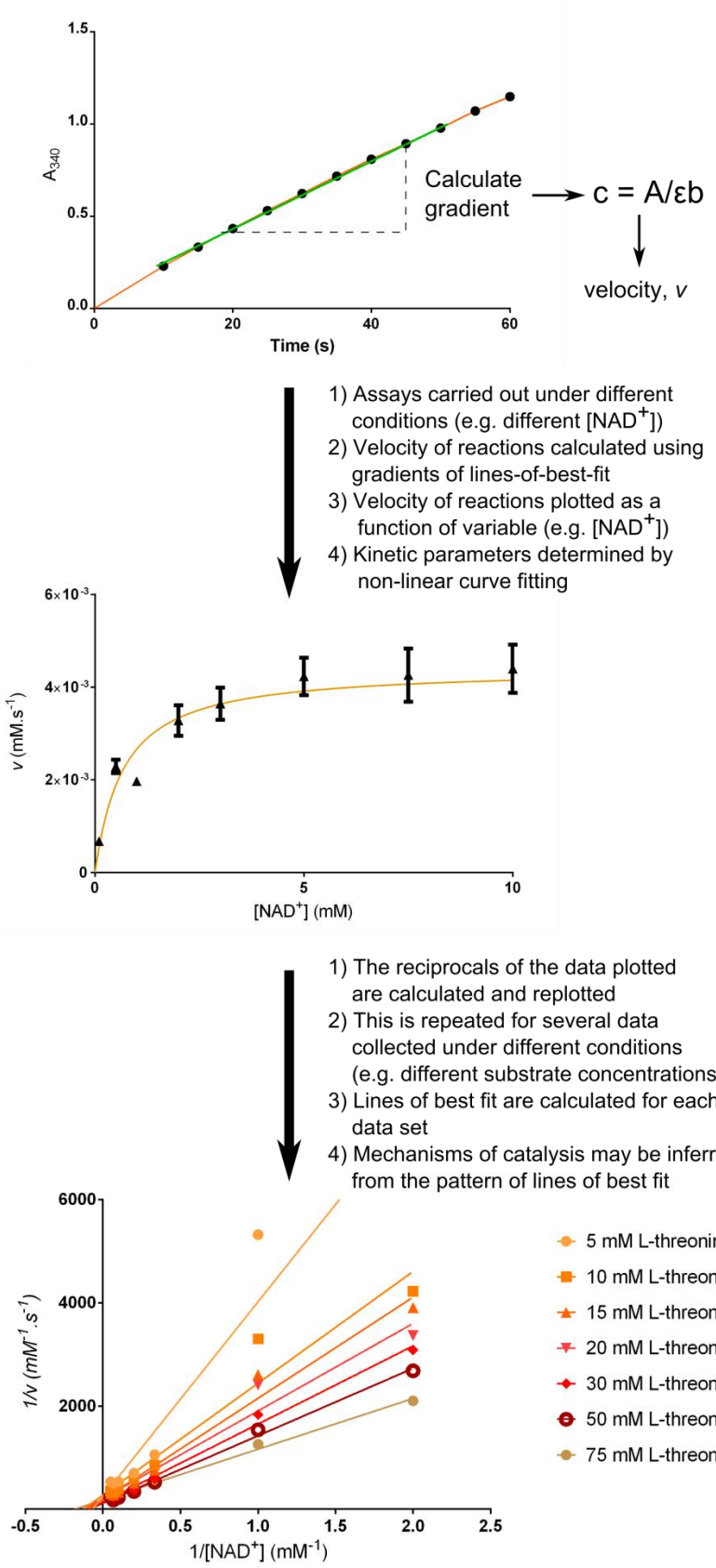


Figure 2.2.5 - scheme illustrating the derivation of kinetic parameters from data collected in enzyme kinetic assays. Rates of reaction (*v*) calculated from raw data (top) were used to construct plots of *v* as a function of [substrate] (middle), from which non-linear curve fitting was used to calculate the parameters K_M and V_{max}. The reciprocals of *v* were plotted against the reciprocals of [substrate] in Lineweaver-Burk plots (bottom).

Recombinant TDH from the bacterium *Clostridium difficile* was expressed and purified by Professor Jon Cooper and Dr Peter Erskine at the Laboratory for Protein Crystallography. In order to determine its kinetic behaviour, the enzyme was assayed in a similar manner to TDH from *T. brucei*. At 37°C, enzyme activity was assayed at 0.1, 1, 5, 10, 15, 20, 30, 50 and 75mM L-threonine, and at 0.1, 0.5, 1, 3, 5, 10 15 and 20mM NAD^+. The assay was carried out using a FLUOstar Omega plate reader (BMG Labtech) and the data were analysed using the aforementioned software.

2.2.2 Optimisation and Alteration

The assay can be optimised, following the same principles as in assay design (see above). The tests required for optimising assay conditions also offer an opportunity to discover more about the enzyme. As the pH of the reaction buffer, the presence of ions and the concentration of the enzyme can all influence enzyme activity, these factors were altered to determine such effects. At saturating concentrations of L-threonine and NAD^+, the rate of TDH activity was measured as previously described at pH 6.5, 7.5, 8.5, 9.5, 10.5, 11.5 and 12.5. To measure the effect of enzyme concentration and to determine if the relationship between it and activity was linear, the activity of a series of two-fold dilutions of TDH was measured at saturating substrate concentrations. See below for a description of assays carried out to test the effect of numerous monovalent and divalent metal ions on TDH activity.

As the testing of compounds for inhibitory activity against TDH would possibly involve the dissolution of compounds in DMSO, the effect of DMSO concentration of TDH activity was investigated. DMSO concentrations of 0-20% were tested in combination with saturating concentrations of substrates, and the corresponding rates of reaction were compared.

To investigate the possibility that the His-tag on TDH would affect its catalytic activity, the tag was removed, following the same protocol as described above (see pull-down assay). Rates of reaction for His-tagged TDH and cleaved TDH were measured across ranges of [NAD^+] and [L-threonine].

2.2.3 Mode of Inhibition studies (MOI)

To investigate the modes of inhibition of TDH, studies to characterise known inhibitors of TDH were carried out. This was done with pyruvate(124,197), L-*allo*-threonine(119,198–201), methylglyoxal(202) and tetraethylthiuram disulphide (TETD or disulfiram)(138,139), all of which are reported in the scientific literature to inhibit a TDH enzyme. For pyruvate, L-*allo*-threonine and methylglyoxal, stock solutions were made up in reaction buffer, and they were added to the reaction mixture, maintaining a final volume of 980µl, prior to reaction initiation. Guided by published inhibitory concentrations, a range of inhibitor concentrations were tested against TDH in the presence of saturating concentrations of L-threonine and NAD^+. The resulting concentration-inhibition curve was used to inform on the inhibitor ranges to be used in further inhibitory assays.

To gain suitable data on the mode of inhibition (MOI) of a particular inhibitor, it is necessary to test the effect that a range of inhibitor concentrations has on the curve produced by also testing a range of substrate concentrations and measuring the rate of activity. In this way, a fixed concentration of the inhibitor and a fixed concentration of one substrate were tested with a range of concentrations for one of the substrates. The *v* versus [substrate] curves gained by repeating this over a range of inhibitor concentrations are then fit to equations describing different MOIs by multiple non-linear least squares regression. The different modes of inhibition are described in Figure 2.2.6.

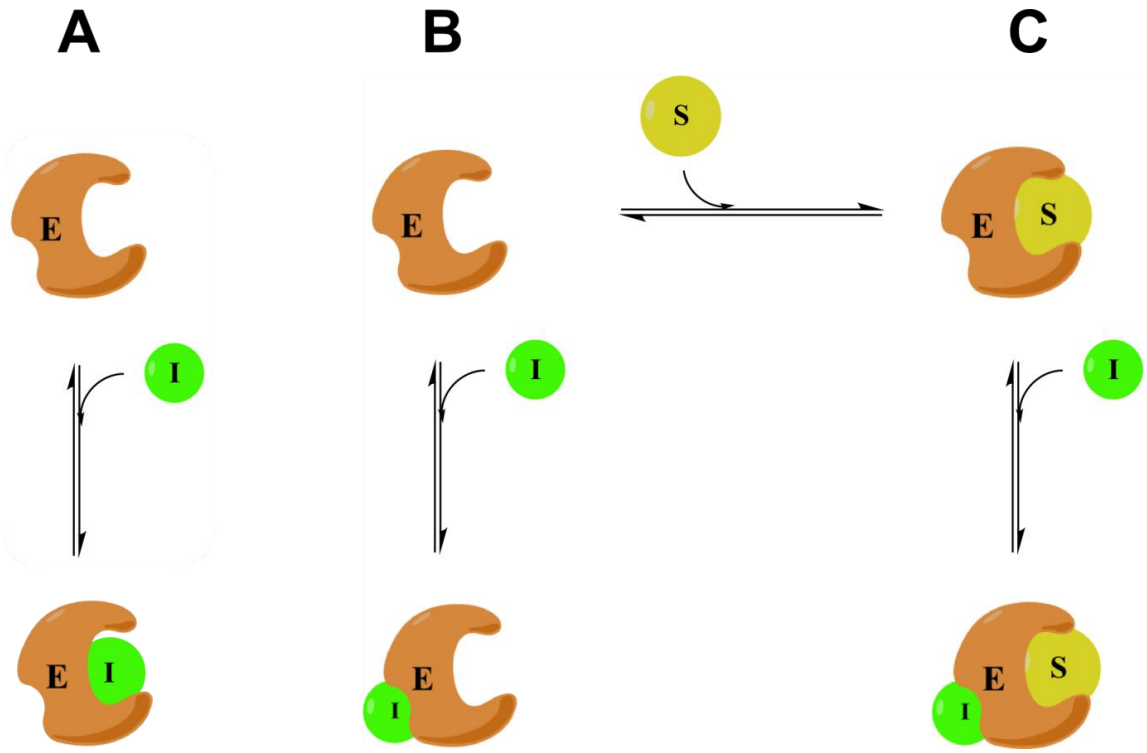


Figure 2.2.6 - modes of enzyme inhibition. A - competitive inhibition, where the inhibitor, I has affinity for the free enzyme, E, and competes with the substrate for binding. B and C – noncompetitive inhibition, where the inhibitor has affinity for both the free enzyme and the enzyme substrate complex. (Note: such inhibitors do not necessarily have to bind to allosteric sites, as suggested in the diagram). C – uncompetitive inhibition, where the inhibitor has more affinity for the enzyme-substrate complex.

Prior to data analysis, the data were normalised by setting the mean maximum velocity of catalysis to a value of 1.0 and by calculating all other velocities as a fraction of this. Then, the analytical software Prism (Version 6, 2013, GraphPad Software, http://graphpad.com) was used for the fitting of the data to the equations describing:

- Competitive inhibition:

$$v = \frac{V_{max}[S]}{[\mathrm{S}] + K_M\left(1 + \frac{[I]}{K_\mathrm{i}}\right)} \qquad \textbf{13}$$

- Uncompetitive inhibition:

$$v = \frac{V_{max}[S]}{[\mathrm{S}]\left(1 + \frac{[I]}{\alpha K_\mathrm{i}}\right) + K_M} \qquad \textbf{14}$$

- Mixed or noncompetitive inhibition:

$$v = \frac{V_{max}[S]}{[S]\left(1 + \frac{[I]}{\alpha K_i}\right) + K_M\left(1 + \frac{[I]}{K_\mathrm{i}}\right)} \quad \textbf{15}$$

K_i is the inhibitory binding constant. Here, the parameter alpha (α) can be used to calculate the affinity of the inhibitor for the free enzyme or enzyme substrate complex(203). An α value of 1 represents a situation where the inhibitor has equal affinity for the free enzyme and the enzyme substrate complex. The R squared value, which is a figure between 0 and 1 describing the fit of the data to each equation, was used to determine the most probable MOI, as indicated by the data. Assays to establish IC_{50} were carried out using the Ultrospec 3000 spectrophotometer. Because of the large number of assays required to gather enough data for MOI determination, those assays were carried out using the FLUOstar Omega plate reader. Solutions of the relevant combinations of L-threonine, NAD^+ and reaction buffer were made up to a volume of 190µl in wells on a clear, plastic, flat-bottomed 96 well plate and incubated at 37°C for least five minutes prior to the initiation of the reaction. The enzyme was stored at a concentration of 0.10 – 0.33 mg/ml (2.6-8.7 µM) in storage buffer and 10µl was injected into the solution to initiate the reaction. Absorbance measurements were made at 320, 340, 400 and 550nm at five second intervals over a two minute period. Measurements made between 10 and 50 seconds inclusive, were used to calculate the catalytic rate, as described above. The plate reader automated this process for all assays set up in a plate.

As TETD showed evidence of time-dependency, the assay had to be altered to allow pre-incubation of the inhibitor with the enzyme, prior to the initiation of the reaction. In these assays, the degree of inhibition was measured as a function of time to calculate the parameter k_{obs}, which is the rate constant for time-dependent enzyme inhibition.

$$k_{obs} = \frac{-\left(\ln\left(\frac{v}{v_0}\right)\right)}{t} \quad \textbf{16}$$

v is the reaction velocity at time, *t*, and v_0 is the reaction velocity at $t = 0$. By measuring k_{obs} at a number of inhibitor concentrations, it was possible to calculate inhibitory binding constants. MOI studies were carried out by determining k_{obs} at a fixed concentration of TETD, but at a range of L-threonine and NAD^+ concentrations. The data were then fit to equations describing:

- Competitive inhibition:

$$k_{obs} = \frac{k}{1 + \frac{[S]}{K_M}} \quad \textbf{17}$$

- Uncompetitive inhibition:

$$k_{obs} = \frac{k}{1 + \frac{K_M}{[S]}} \quad \textbf{18}$$

- Noncompetitive inhibition (α = 1):

$$k_{obs} = k \quad \textbf{19}$$

In these equations, *k* is treated as an empirical value for curve fitting purposes(194). As before, the most probable MOI was determined by observing the R squared value. These assays were carried out in the plate reader in a similar manner to the other MOI assays. However, TETD, which was dissolved in DMSO, was added to the enzyme solution in the appropriate concentration. Every well on a particular plate contained fixed concentrations of substrates. The reaction was initiated as above with the addition of 10µl of enzyme solution (pre-incubated with TETD). A control well, where the reaction was initiated with enzyme solution pre-incubated with an equivalent volume of DMSO was measured directly after each experimental well, so that the fractional activity or percentage inhibition could be determined by comparing the activity between each pair of wells. The data were fit to the relevant equations using Prism. The equations for time-dependent inhibition were built as custom equations, as they were not pre-installed. In the case that any of the data from MOI studies showed evidence of cooperativity of the inhibitor, modifications to the equations were made by applying a Hill coefficient to [I].

For some enzymes, high concentrations of substrate can also inhibit the enzyme, either by effects of the substrate itself, or by effects of the product. To investigate substrate inhibition, another equation within the Prism software was used to model the phenomenon:

$$v = \frac{V_{max}[S]}{K_M + [S]\left(1 + \frac{[S]}{K_\mathrm{i}}\right)} \qquad \textbf{20}$$

The equation is able to determine both the K_M of enzyme activity and the K_i of enzyme inhibition.

2.2.4 Further studies of enzyme activity

The effect of naturally-occurring amino acids on TDH activity was investigated to explore the possibility that TDH is not fully selective for L-threonine and to see if any other amino acids inhibited TDH. The following amino acids were tested in duplicate assays at 60mM concentrations in the presence of saturating concentrations of L-threonine and NAD^+: L-alanine, L-cysteine, L-aspartic acid, L-glutamic acid, L-phenylalanine, glycine, L-histidine, L-isoleucine, L-lysine, L-leucine, L-methionine, L-asparagine, L-proline, L-glutamine, L-arginine, L-serine and L-tryptophan. It was not possible to test L-tyrosine in this assay, due to its low solubility. These assays were carried out according to the protocol used for the MOI assays described above. A second set of assays were carried out in the same way, but L-threonine was excluded to investigate the possibility that TDH can catabolise other amino acids.

The effect of different monovalent and divalent cations on TDH activity was investigated using assays carried out at saturating substrate concentrations. The following compounds were tested at 1mM by incubating with TDH in storage buffer overnight: Na_2SO_4, $Zn(C_2H_3O_2)_2$, $ZnSO_4$, $CoCl_2$, $MnCl_2$, $CdCl_2$, $CuCl_2$, $FeSO_4$, $FeCl_3$, $CaCl_2$. They were then tested by comparing with the activity produced by a DMSO control. Four assays were carried out for each condition and mean values of the relative activity were calculated. Using the same assay, the following

compounds were tested in duplicate at 0.25mM, 0.5mM, 1mM, 2mM, 5mM and 10mM for their effect on TDH activity: KCl, RbCl, CsCl and NH_4Cl.

Following some interesting observations, more assays were carried out to determine the effect of reaction constituents on the rate of activity and the stability of TDH over time. In these assays, TDH was pre-incubated in reaction buffer, with and without DMSO, for different lengths of time, prior to initiating the reaction with the addition of substrates. The concentration of glycerol and the addition of salts, NaCl, KCl and RbCl, were investigated in this respect. Then further assays were carried out to measure the effect of KCl and RbCl in the reaction buffer on the relationship between [L-threonine] or [NAD^+] and rate.

The following compounds were tested for time-dependent inhibition, using the same protocol as for the measurement of time-dependent inhibition of TETD: 1mM $AgNO_3$, 1mM $ZnSO_4$ and 1mM $HgCl_2$. The following compounds were also tested for inhibition of TDH, either by incubation overnight or using a time-dependent assay as described above: N-ethylmaleimide (0.5mM, time-dependent assay), iodoacetamide (5mM, time-dependent assay) and EDTA (5mM overnight).

2.3 Virtual Screening

The process of identifying of new active compounds during screening can be divided into three steps. The first step, hit generation involves the identification of active molecules in a primary screening assay. Secondly, hit confirmation is carried out in a secondary assay to confirm that the active compounds (hits) generated are indeed active. Finally, hit validation is performed to verify the identity of the compound exhibiting activity, and to confirm that the activity is specific and not due to other factors, such as interference with the assay signal.

Virtual screening formed a part of the hit generation strategy for the overall screening process presented in this thesis. However, to effectively make use of virtual screening, it was necessary to carry out the three steps of hit generation, hit confirmation and hit validation.

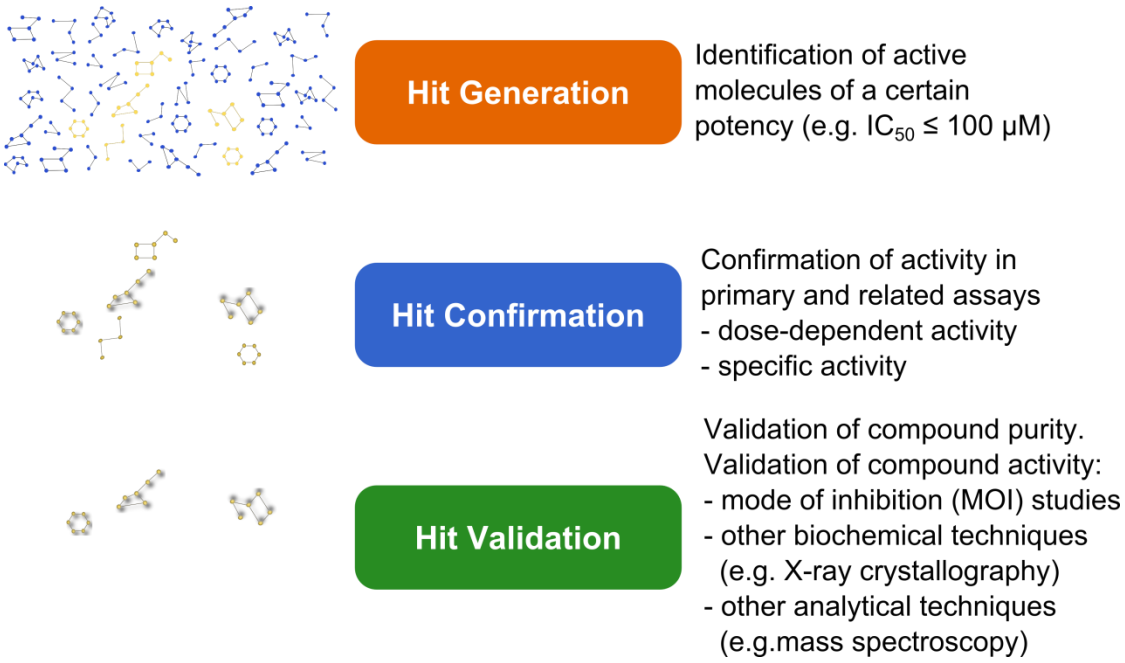


Figure 2.3.1 - the identification of active molecules, 'hits', in drug discovery. The process illustrated applies to both virtual and in vitro screening.

2.3.1 Docking Software and AutoDock 4

Virtual screening is a powerful tool in drug discovery that is made readily-accessible by the wide availability of docking simulation software to academic and private-sector researchers. Docking programs predict the strengths of interactions between different molecules, commonly between small molecules and proteins.

AutoDock is the most cited of all the available docking simulation software packages(29). AutoDock uses information on atom-types, partial charges and the electrostatic potential of different atoms to predict the interaction between ligands and macromolecules (called the 'receptor' in AutoDock). It makes use of file formats called PDBQT (.pdbqt) files for both ligands and receptors that are similar to the .pdb files that are familiar to protein crystallographers, except they include additional information on AutoDock4 atom types, partial charges (Q) and rotatable bonds (T). It simulates docking by the construction of a grid, which maps the possible interactions between all the different atom types on a 'receptor' molecule and atom types on a ligand molecule. These grid maps are then used to calculate binding energies between a ligand and the atom types within the grid. The program AutoGrid performs the calculation of grid maps and AutoDock performs the docking predictions. They can be run entirely from the command line, but the GUI program AutoDockTools (ADT) was used to prepare the receptor and search grid, and to visualise some docking results.

AutoDock can predict binding interactions by different methods: use of a genetic algorithm, by simulated annealing and by a local search. The most effective method available in AutoDock 4 is the combined genetic algorithm-local search (GA-LS) method, termed the "Lamarckian" genetic algorithm. The GA-LS is based on the Lamarckian theory of evolution, which proposed

that phenotypic characteristics developed during an organism's lifetime were passed on to their offspring(204). AutoDock generates a population of ligand conformers, which undergo crossover and mutation of "genotypic" conformational characteristics before a local conformational search for energy minima is performed. The "fittest offspring" – those conformers with the lowest internal energy – become parent conformers for future generations, which undergo a repeat of this cycle for a user-specified number of times. Eventually one conformer with the lowest energy remains at the end of a docking run. This conformer is then tested for the strength of interaction with the receptor which is calculated by combining the intramolecular energy of the ligand and the intermolecular energy of the ligand-receptor complex. These calculations include terms that relate to the changes in Van der Waals forces, hydrogen bonds, electrostatics and desolvation that occur on binding(24,205). AutoDock version 4.2 was used in this investigation. The recommendations in the AutoDock 4 tutorial(206) were used to guide the design of this experiment.

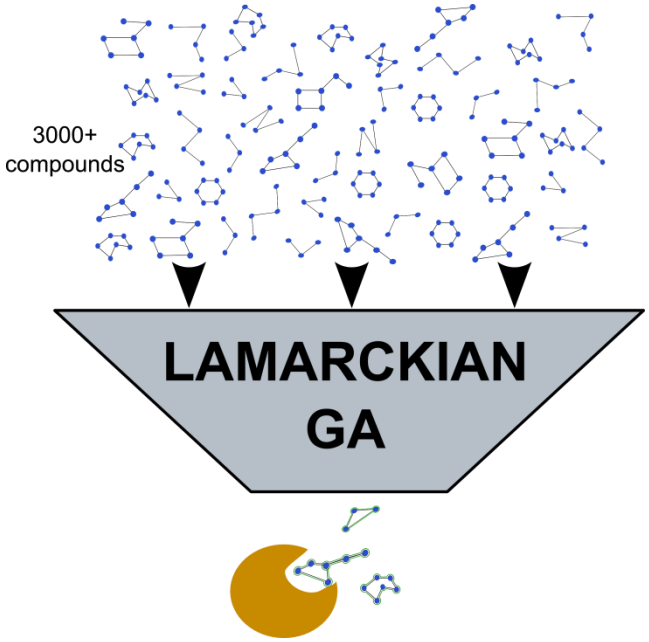


Figure 2.3.2 - schematic of virtual screening using AutoDock. Stick and ball figures represent conformations of ligands. More than 3000 compounds were screened in the virtual screening campaign. The best conformations of those ligands were selected by the Lamarckian Genetic Algorithm and the interactions between these conformers and the target were simulated.

2.3.2 Screen Preparation

The importance of taking into consideration receptor or protein conformational flexibility in the design of virtual screening experiments is now widely recognised. As TDH binds a cofactor, and was found to exhibit conformational flexibility, four virtual screens were planned using a different receptor each:

- Virtual Screen 1 – NAD^+-bound TDH in conformation A
- Virtual Screen 2 – NAD^+-bound TDH in conformation B
- Virtual Screen 3 – Unbound TDH in conformation A
- Virtual Screen 4 – Unbound TDH in conformation B

Structures were taken from PDB files of TDH structures solved in this study. If a PDB file contained several TDH monomers, one would be chosen on the basis of it having the best fit to the $2F_o$-F_c electron density map. For Virtual Screen 1, 2 and 3, the receptor file was cut from the text in PDB files. For Virtual Screen 4, the same data as Virtual Screen 2 was used, but the NAD^+ molecule was removed from the file.

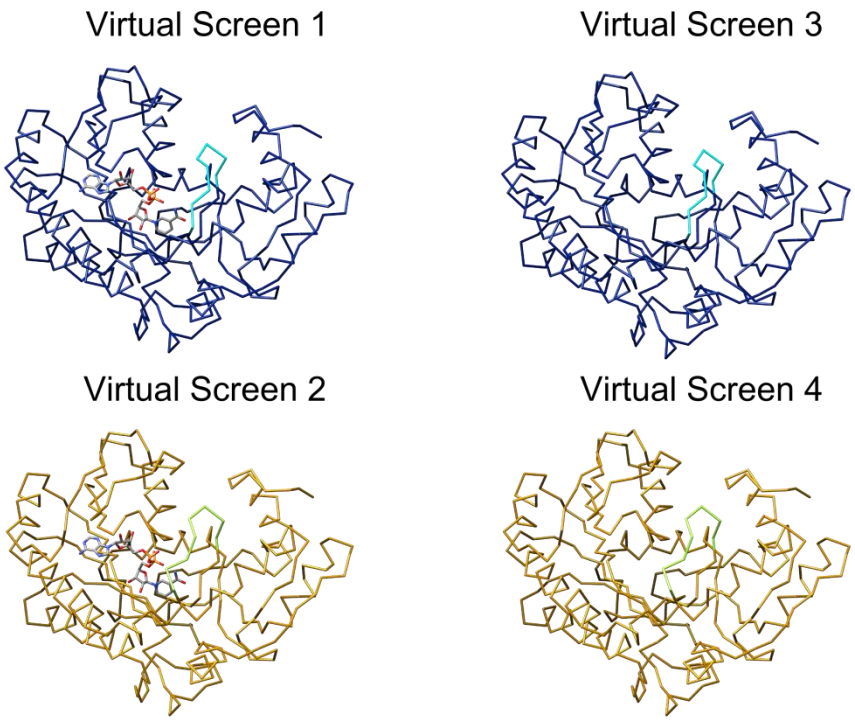


Figure 2.3.3 - wire representations of the four TDH models used as receptors for virtual screening with AutoDock. Blue structures represent TDH in its 'open' conformation and yellow representations show TDH in its 'closed' representation.

Ligands and macromolecules can be converted to the necessary PDBQT files using the GUI, ADT. First, hydrogen atoms are added wherever appropriate. Then partial charges (Gasteiger charges) are computed for each item. Then, non-polar hydrogens are merged (removed), as they are assumed not to affect charges, leaving just polar hydrogen atoms. For receptor molecules, AutoDock4, is able to define amino acid side chains as flexible(24). For this, a second PDBQT file needs to be written and this will contain information on the torsions, which are the rotatable bonds, of the side chain atoms.

As stated earlier, AutoDock is able to predict intermolecular binding energies by calculating a map that carries information on all the atom types and charges in a specified area. These maps are calculated across a grid that defines the search area for intermolecular interactions. A cubed search grid of 60 points at 0.375Å spacing along each axis was set. The C4 atom of the nicotinamide ring of NAD^+, was chosen as the centre of the grid as this is believed to be the site of the TDH reaction(124). When NAD^+ was not present in the model, the oxygen atom of the

hydroxyl group of the nearby tyrosine residue (Tyr144, see section 3.1) was used to centre the grid. Atom types were set 'directly' by entering a list of all atom types found in the compound libraries. The program AutoGrid was run using these parameters to determine the map types to be used in docking.

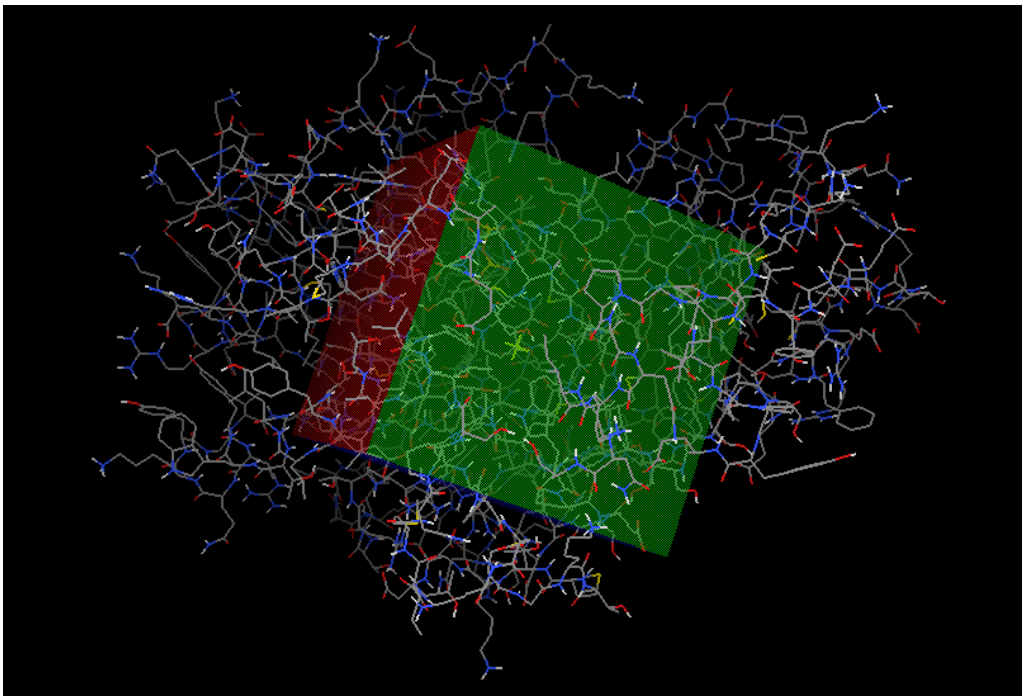


Figure 2.3.4 - the assignment of the search grid (search area) in ADT. The yellow cross shows the centre of the grid, which is the C4 atom of the nicotinamide ring of NAD^+.

Once it was decided which TDH conformers would serve as receptors, it was important to ensure that each TDH model to be used was likely to give accurate docking predictions. Based on the study of a similar TDH enzyme(124), the residues present in the active site in a *T. brucei* TDH model, and the result of a docking simulation with L-threonine, five active site residues were identified as being important in binding: Ser82, Thr119, Met81, Tyr144 and Thr186. A number of docking simulations were carried out using default GA-LS parameters with different combinations of flexible residues and a completely rigid TDH model (designated for 'Virtual Screen 1'; see above), using L-threonine, glycerol or pyruvate as a ligand. The validity of the 'receptor' model was judged based on the plausibility of the binding position for L-threonine, the predicted binding energies and the clustering of binding poses. A completely rigid molecule was chosen for all screens due to the more plausible binding energies predicted for all ligands (see results section for more detail on this).

A second series of experiments aimed to determine the most time-efficient parameters to run AutoDock with. The AutoDock developers recommend that at least 50 docking runs are performed per ligand(207) so this value and 100 and 150 docking runs were compared. The majority of the ligands in the compound library had less than six rotatable bonds, and it is recommended that a limit of between 250000 and 25 million energy evaluations is set for

ligands with 1-10 rotatable bonds, so 250000, 500000, 1.25million, 2.5million and 25million were tested. Finally, initial genetic algorithm populations of 100, 200 and 300 were tested. 50 docking runs and 1.25million energy evaluations were chosen as higher numbers had an insignificant effect on docking results but a detrimental effect on the simulation time. A genetic algorithm population of 300 was chosen based on the assertion of Hetényi *et al.*(208) that a population of 300 produces better results and the fact that it does not increase simulation time significantly. Other parameters were left as default settings.

2.3.3 Compound Libraries

In deciding which compounds to screen against TDH, several strategies were considered to increase the likelihood of identifying active compounds or 'hits'. Principally, these strategies aimed to increase the 'diversity' of the library by sampling a larger area of chemical space, to identify potential hits by exploiting knowledge of the structure of TDH and its ligands, and to increase the probability of finding hits by enriching the library with molecules that are known to have some bioactivity, such as natural products or known trypanocidal drugs. With this in mind, the library used in the virtual screening can be classified as follows:

- Diversity (NCI Diversity Set II) – A library of compounds that is designed for chemical diversity contains a large range of different chemical structures.
- Ligand analogues – A library built from compounds that shared some structural similarity with known TDH ligands, glycerol, pyruvate and L-threonine.
- Natural – A library of products from natural sources, where either the natural source (e.g. a medicinal plant) or the compound itself had published evidence of bioactivity. This library was also constructed to produce diversity in the library.
- Knowledge – A library of compounds that were identified based on knowledge of their: trypanocidal activity, structural similarity to a trypanocidal compound, published activity against TDH from another species, or role as a ligand of enzymes sharing similar 3D structures with TDH.

These libraries and how they were constructed will be discussed in more detail, below.

NCI Diversity Set II

The Diversity Set II of the US National Cancer Institute (NCI) is a library of compounds designed to represent the entire chemical diversity of the compounds held by the institute(209). Compounds were selected to represent a diverse range of different pharmacophores (the combination of molecular features that are responsible for a ligand's action)(209). ZINC is a free online database that currently holds more than 14 million purchasable compounds in a variety of chemical structure file formats(210). The NCI diversity set II deposited in this database, classified as 'usual', which contains structures for all the tautomeric forms of the compounds across a pH of 5 to 9, was downloaded as .mol2 files. There were 2044 files in total.

Ligand analogues

It was thought that structures similar to known TDH ligands would be more likely to produce hits in a virtual screen. A search for analogues of or similar molecules to L-threonine, glycerol and pyruvate was conducted in the ZINC database as follows: an 80% similarity search based on the ligand, limiting results to those with less than 10 rotatable bonds (this was done based on published information citing decreased accuracy of AutoDock predictions with molecules possessing more than 10 rotatable bonds(24)). The top 250 results from each search were downloaded as .mol2 files regardless of whether they were purchasable or not. This was done because the primary concern at this stage was to identify hits and possibly some structure-activity relationships (SARs); it has been suggested that a series within a compound library should contain at least 200 members in order to deduce SARs(211). 750 unique structures were selected, totalling 796 structures once different tautomers were included.

Natural Products

Natural products have served as sources of new drugs or inspiration for new drugs for a long time(212–214). Although natural products are often said to lack "drug-like" properties, they represent an area of chemical space that isn't as well represented by synthetic drugs(215–218) so they can be useful in identifying pharmacophores that otherwise would not be tested. With this in mind, a library of 426 natural products was built by searching for natural products with potential or reported trypanocidal activity. Using the Ovid SP server, a search was conducted in Embase using a combination of the Emtree terms "herb" OR "medicinal plant" with "Trypanosomiasis" executed with an AND operator. This search was later expanded to include non-plant sources by searching for the keywords "natural product" AND "trypano*". Then, compound names were taken directly from literature reports describing activity for the compound or its biological source, or from literature describing the constituents of plants reported to have activity against trypanosomes. One journal paper was found which had applied this approach for anti-trypanosomal drug discovery(219). In that work, Ogungbe and Setzer screened compounds with reported trypanocidal activity against a number of promising drug targets and were able to discover a significant number of hits. The structures of the natural products were obtained as Simplified Molecular Input Line-Entry System (SMILES) string (220–223) representations, which were obtained from the PubChem(224,225) database or by drawing the published structures in the Molinspiration(226) program (which can convert drawn structures to SMILES) embedded in the ZINC website. A total of 426 unique molecules were obtained in this library. It was noted that several of the natural products contained saturated rings. AutoDock has previously been limited by the inability to perform energy minimisations on flexible ring structures(205). Although AutoDock 4 has found a way around this with the introduction of 'glue' atoms, the incorporation of these atom types would be too laborious a task for hundreds of molecules. Openbabel, an open-source chemistry program that can generate, filter, convert and analyse chemical structure files(227,228), was used to select molecules with flexible rings. Then the program Balloon, which generates multiple

conformers of any structure by use of modifications directed by a genetic algorithm(229), was used to generate multiple conformers of these structures – a common strategy used for the virtual screening of ring-containing compounds(230).

Knowledge

Despite being the smallest of the four libraries classified here, this library includes a combination of several hit identification strategies. The first strategy was to build a small collection of compounds with documented trypanocidal activity. Many of the trypanocidal drugs used historically were developed empirically and therefore, have an unknown or poorly characterised mechanism of action(231). Where a mechanism of action is known, there is still the possibility that the drug may act on multiple intracellular targets(37,84). Therefore, a small collection of 21 compounds historically or currently used to treat trypanosomiasis was constructed. Many of the molecules were identified through Steverding's informative historical review of treatments for HAT(231). Twenty-one compounds were downloaded as SMILES strings obtained in the same manner as the natural products library.

A small selection of compounds, obtained from Dr Stephen Hilton at UCL School of Pharmacy (SoP), were found to inhibit TDH enzyme activity by a summer research student working in this lab. As these compounds had structures analogous to TETD, they were included in the virtual screen. PDB files of these compounds were generated using the ProDRG online server(180).

In 2011, it was reported by Alexander et al. that a series of quinazolinecarboxamide (Qc) compounds were able to potently inhibit TDH from mouse embryonic stem cells(136). Using the reported structures, PDB files were prepared in the same way as for the SoP compounds.

PSSC

The technique of Protein Structure Similarity Clustering (PSSC) was originally created by Koch and colleagues to identify scaffolds for the development of compound libraries for high-throughput screening. It stemmed from the observation that, despite the high degree of diversity in the primary structures (amino acid sequences) of proteins in nature, many proteins converge on a smaller set of tertiary structures than one would expect from this diversity. Thus, proteins with low amino acid sequence similarity can share a high degree of 3D structure similarity(232). Following this principle, a search for proteins with similar 3D structures to TDH was carried out. A collection of compounds made from ligands of those proteins was then constructed. The original PSSC approach searched for proteins with up to 20% amino acid sequence similarity with the target, and that had a root-mean-square deviation (RMSD) of less than 4-5Å when the catalytic cores of both protein structures were superimposed. Results were also checked visually for structural similarity. In a similar manner, proteins sharing similar 3D structures with TDH were sought using the same parameters, but with some additional parameters to focus the results. Searches were carried out in two rounds: first using the catalytic domain of TDH, and second using an entire TDH monomer. A search using the

catalytic domain was conducted initially to exclude the large number of results that would be found within the protein's superfamily. The second round, using the full structure of a TDH monomer was subsequently carried out to find any proteins that may have been unnecessarily excluded by following the first strategy. The search was executed on the PDBeFold(233,234) search engine, searching the Structural Classification of Proteins (SCOP)(235) and the entire PDB databases(236,237). The PDBeFold search engine outputs a number of parameters describing the match between the query protein and a query result. These included the percentage sequence similarity, the RMSD of superimposed alpha carbons (Cα) and the Q-score, which gives a measure of structural similarity, whilst taking into account the number of aligned amino acid residues. The P-score represents the probability that a better alignment could have been found by chance, and hence represents the validity of any match. Therefore, all results with a P-score lower than 3 were rejected. Once a protein meeting the criteria for PSSC had been identified, the model was downloaded and superimposed on the query TDH monomer using the 'Matchmaker' tool in UCSF Chimera(182). The 'match→align' tool was then used to calculate and confirm RMSD values. Finally, the aligned proteins were checked visually, particularly in the active site region. For proteins found in the second round of the PSSC search, the proteins were matched and aligned with the catalytic domain of TDH, and RMSD values were calculated for those matched regions, alone.

Ligands and inhibitors of all proteins found by PSSC were found primarily through the online enzyme information database, BRENDA (BRaunschweig ENzyme Database)(238), as all the proteins identified were enzymes. Additional inhibitors were found by searching the enzyme name and the word "inhibitor" on the search engine Google(239). Further information was also found from the Binding database(240). Structures of the reported ligands were sought from the PubChem(224) and ChEMBL databases, and compounds were downloaded as SMILES strings, where available. Where SMILES strings were not readily available, published structures were re-drawn in the chemical drawing software Marvin Sketch (version 6.0.2, 2013, ChemAxon, (http://www.chemaxon.com)) and converted to SMILES.

All SMILES were converted to .pdb or .mol2 file formats by Open babel, using the "-gen3d" option to generate stereospecific 3D coordinates in the case of SMILES. Open babel was also used to split multi-molecule files. As AutoDock 4 does not include parameters for arsenic atoms, they were replaced with phosphorus atoms in melarsoprol, melarsen and melarsen oxide. This strategy has been used previously to mimic arsenic(241). All .pdb and .mol2 files were converted to the appropriate format for AutoDock, PDBQT, by use of the 'prepare_ligand4.py' script supplied by the Scripps laboratory. To examine the atom types represented by the entire compound library, the 'examine_ligand_dict.py' script was invoked, outputting a file containing all atom types.

2.3.4 Screen Execution and Analysis

To carry out the screening of such a large library of compounds, each individual ligand was organised into a folder that was linked to a common folder containing the 'receptor files' (grid maps, PDBQT files). A Docking Parameter File (.dpf), which contained the settings for each screen was generated and contained in the folder for each ligand. The screen was then executed by use of a c-shell script that executed each screen sequentially, according to the parameters set in the .dpf file. Alternatively, the screen was executed across a cluster of four computers, managed by the Sun Grid Engine, with four screening experiments running simultaneously.

A script provided with the AutoDock4 tutorial was modified to automate the summary and ranking of the results of virtual screening. The results of all 50 docking experiments for each ligand were clustered into groups of results that varied by $\leq$ 2.0 Å RMSD in their atomic coordinates. The cluster of the largest size, with the best docking score (binding energy) was used to report the result for that ligand. Comparing these binding energies, all ligands with a binding energy of -9.50 kcal.mol^{-1} or lower was considered a 'hit'. This threshold was chosen as it was thought that this would select a subset of molecules that scored highly and were therefore more likely to be *in vitro* hits of TDH. Compounds for in vitro testing were validated visually by checking for steric clashes, unnatural conformations and appropriate binding modes using ADT and PyRx. The sizes of the clusters and the variation between different docking results were also used to verify the consistency of docking predictions. Where multiple conformations of a ligand had been represented by separate files, only the highest scoring of all the related files was counted.

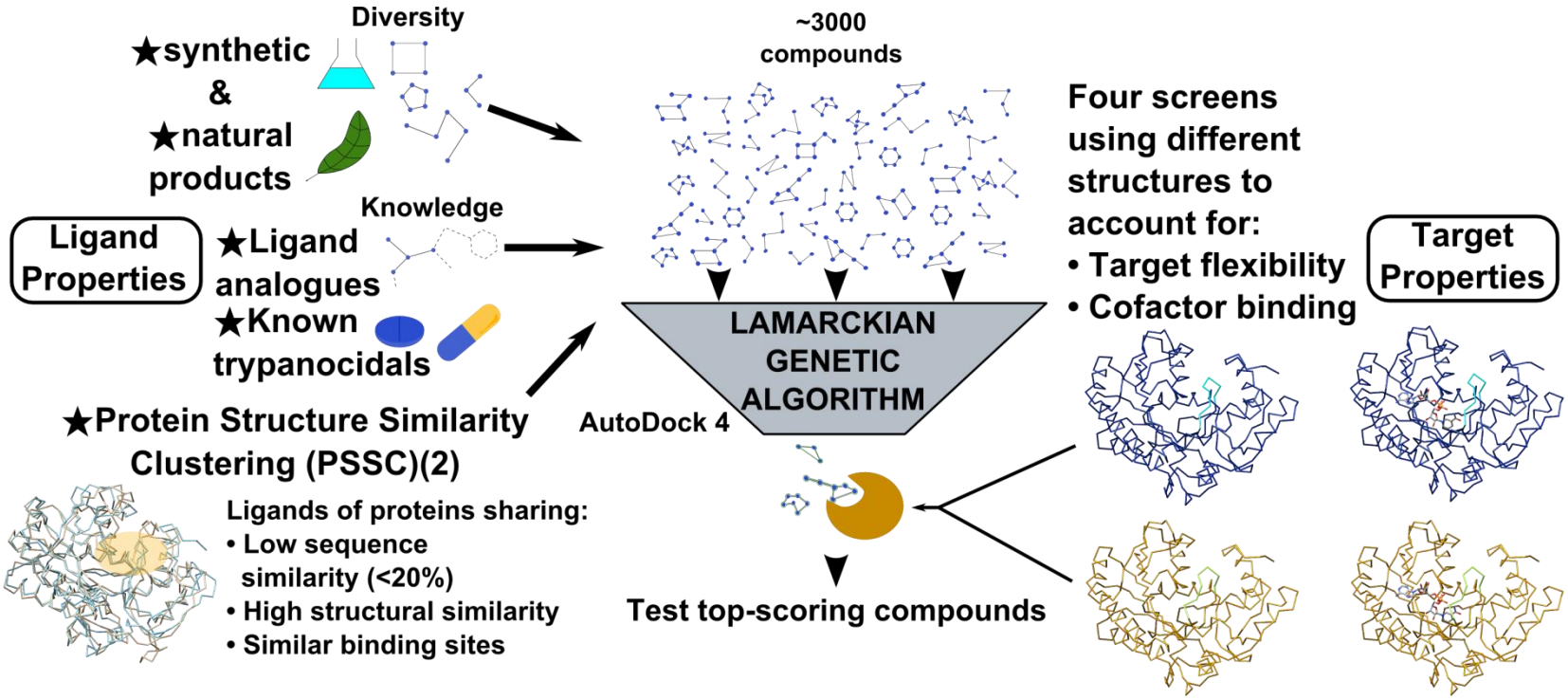


Figure 2.3.5 - Schematic summarising the different strategies employed in the virtual screening campaign against TDH. Ligand properties were taken into account when building compound libraries and different TDH models were used for screening to represent the variation in the target's different conformational and bound states.

2.4 *In vitro* Screening

2.4.1 Libraries

The *in vitro* screening experiments provided another opportunity to approach hit identification with several different strategies, as with the virtual screening experiments.

Fragment-based drug design aims to identify molecules (fragments), which are typically smaller than drug molecules, and have low or moderate affinity for the drug target. Once information on their binding sites are obtained, by X-ray crystallography, for example, fragments binding to different sites can be linked to form a single molecule that can have much higher potency than the original fragments(19). To apply this approach to the discovery of TDH inhibitors, two fragment libraries were screened as part of the *in vitro* screening efforts. A library of 500 fragments was obtained from Maybridge and a library of 358 was obtained from the 3D Fragment Consortium. The Maybridge fragment library is designed for diversity and has been designed for ease of synthesis, later in the drug design process(242). The library designed by the 3D Fragment Consortium was designed to include fragments with more three-dimensional geometries than are typically contained in most commercial fragment libraries(243).

Virtual screening has been shown to be a cost effective way of identifying *in vitro* inhibitors of drug targets. Thus, a selection of the compounds identified by AutoDock as potent TDH inhibitors was compiled for screening in the virtual screening experiment. Several of these compounds were obtained free of charge from the Developmental Therapeutics Program (DTP) of the NCI and US National Institutes of Health (NIH). Others were obtained from commercial vendors. This set of compounds will be referred to as the 'custom library'.

A more traditional approach to hit identification is to screen small molecules with "lead-like" qualities, in the hope that one can discover potent inhibitors. Identification of such a molecule can potentially save time downstream in the drug discovery process, by reducing the amount of optimisation associated with molecules that lack lead-like or drug-like qualities. The Asinex chemical libraries are designed to possess chemical diversity and include compounds with favourable properties related to absorption, distribution, metabolism, excretion and toxicity(244). Compounds from the Asinex library were selected in collaboration with UCL ChemiBank. UCL ChemiBank screened their library of 20,000 Asinex compounds against two TDH structures in a virtual screen using the docking software GOLD. The top 751 results were provided for screening against TDH. A further 249 molecules from the Asinex library were provided by:

- A similarity search for molecules sharing 0.85 structural similarity with hits from the AutoDock virtual screen;
- A substructure search for compounds matching the structures of hits identified by screening of the Maybridge(242) and 3D Fragment Consortium(243) libraries.

These searches were carried out using the chemical database manager Instant JChem (Version 6.1, 2013, Chemaxon(http://www.chemaxon.com)).

2.4.2 Screening Assay & Format

The screening of a large quantity of different molecules required a change in the enzyme assays described above. To increase the chances of identifying inhibitors with competitive, noncompetitive and uncompetitive MOIs, screening was conducted at substrate concentrations that were close to their K_M values or half-saturating concentrations. In this way, competitive inhibitors would not be excluded by testing high substrate concentrations and uncompetitive inhibitors would not be excluded by testing inhibitors at substrate concentrations that were too low. The NAD^+ concentration used was 1.2mM, whilst the L-threonine concentration used was 6mM. A further alteration of the assay was required to be able to detect time-dependent inhibitors. Therefore, the enzyme was pre-incubated with the test compounds and reaction buffer for 30 minutes before initiating the reaction with the addition of a mixture of L-threonine and NAD^+. Assays were prepared in 96- or 384-well plates (Greiner), using a 96-head fluid handler (FluidX). After incubation in the plate reader at 37°C or 25°C for at least 5 minutes, the reaction was initiated and measured automatically using a POLARstar Omega plate reader (BMG Labtech), as described above for the MOI studies. To reduce the overall time taken to read a single plate of assays, the reaction time was shortened: measurements were taken at two-second intervals and the rate of reaction was calculated from the gradient of the line of best fit between measurements taken from 10 to 22 seconds inclusive.

Once the general format had been chosen, the assay was validated using known TDH inhibitors. As there are resource constraints in screening experiments, it is useful to reduce the need for repeat experiments. Validation of a screening assay can increase the confidence that the hits identified will not be false positives and that false negatives will be minimised. The plate uniformity test (PUT) developed by researchers at the NIH and Eli Lilly and Company is a test that provides useful information on the accuracy of an assay and its ability to correctly identify hits(203). The layout of a PUT is shown in Figure 2.4.1.

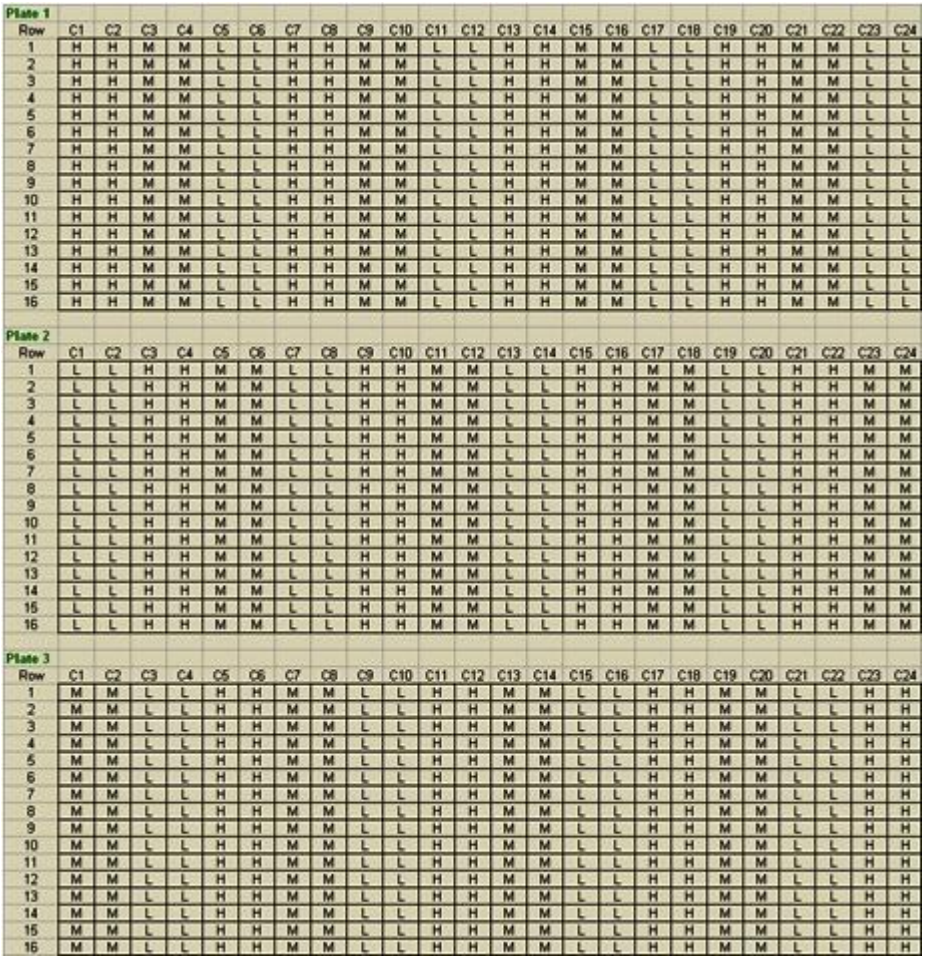


Plate 1																								
Row	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24
1	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
2	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
3	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
4	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
5	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
6	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
7	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
8	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
9	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
10	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
11	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
12	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
13	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
14	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
15	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L
16	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L

Plate 2																								
Row	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24
1	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
2	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
3	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
4	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
5	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
6	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
7	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
8	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
9	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
10	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
11	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
12	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
13	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
14	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
15	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M
16	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M

Plate 3																								
Row	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24
1	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
2	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
3	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
4	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
5	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
6	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
7	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
8	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
9	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
10	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
11	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
12	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
13	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
14	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
15	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H
16	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H	M	M	L	L	H	H

Figure 2.4.1 - The layout for PUT in 384-well plates. H = 'High' signal; TDH was incubated alone with DMSO as a control. M = 'Mid' signal; TDH was incubated with disulfiram or L-*allo*-threonine to produce a signal around 50% of the maximum. Assays in the High and Mid conditions were initiated by addition of substrate. L = 'Low' signal; as for the 'High' well, but the reaction was initiated with buffer and not substrate to reflect the expected signal where there is zero activity.

PUT assays were carried out using 40µM TETD or 7mM L-threonine in the 'Mid' conditions. Substrate was replaced with reaction buffer for the initiation of the assay in the 'Low' condition. TETD was replaced with DMSO and L-*allo*-threonine was replaced with reaction buffer in the 'High' condition.

Informed by the results of the PUT assay, the final conditions for the screening assay were established as follows:

- The reaction buffer was changed to include 1mM KCl, in addition to 200mM Tris.HCl at pH 8.5. The presence of KCl in the buffer conferred greater stability of TDH over the time course of measurement of one plate of reactions. This became the ordinary reaction buffer for use in future TDH assays.
- The storage buffer for TDH was also changed to include 100mM KCl, in place of 100mM NaCl.
- Reactions were prepared in 96 or 384-well plate format, with a final volume of 100µl.
- 10µl of 7.9 µM (0.3mg/ml) TDH was pre-incubated at room temperature for at least 25 minutes with the test compound and reaction buffer, making a volume of 82µl. The

reaction mixture was mixed by five cycles of aspiration and dispensing with the FluidX liquid handler before pre-incubation.
- After pre-incubation, the plate was further incubated for five minutes at 25°C in the plate reader.
- Reactions were initiated by addition of 18µl substrate solution and measured at two-minute intervals over 22 seconds automatically by the plate reader as described previously.

All test compounds were stored in stock solutions of 10mM in DMSO, except for one compound, 2-{[3-(4-bromophenyl)-3-oxopropyl]amino}-3-hydroxybutanoic acid (BPOB), which required dissolution with DMSO and 1% v/v 880-ammonia. Fragments were tested for inhibition in duplicate, at 500µM. Compounds in the custom library were tested once at 500µM and once at 100µM. All Asinex compounds were tested in duplicate at 100µM. Each row in each plate contained a control well that contained DMSO in place of a test compound.

Hits were identified using the following criteria:

- Fragments – exhibiting $\geq$ 50% inhibition (relative to control) at 500µM in two separate assays.
- Custom Library – exhibiting $\geq$ 50% inhibition at 500µM and significant inhibition at 100µM.
- Asinex – exhibiting $\geq$ 50% inhibition at 100µM in two separate assays.

For hit confirmation of all the observed hits, two-, three- or four-fold dilutions were prepared by mixing with DMSO using the FluidX liquid handler. Compounds were then tested for concentration-dependent inhibition of TDH. Compounds that failed to show this effect, or whose activity disappeared below the original test concentration, were rejected.

Hit validation aims to establish that all hits exhibited a specific MOI and that their inhibition is not due to interference with the assay or non-specific denaturation of the enzyme. One way in which compounds may appear as hits in a high-throughput assay is by self-aggregation to form colloids, which inhibit enzyme activity in a non-specific way. This mode of inhibition has been well described by Feng, Shoichet and colleagues, who developed an assay that added detergent to the reaction buffer to prevent colloid formation, thereby abolishing the activity of "promiscuous" non-specific inhibitors(245,246). Similarly, a reaction buffer that contained the detergent Triton X-100 at a concentration of 0.01% v/v was prepared for these assays. L-threonine and NAD^+ stock solutions were then prepared with this reaction buffer. Fragments were tested in the primary assay conditions at their IC_{50} against a DMSO control in duplicate: once in the ordinary reaction buffer, and again in the reaction buffer with detergent. Compounds from the custom library were tested in a similar manner, at concentrations that had showed inhibition during hit confirmation assays. For the Asinex compounds, the hit confirmation assay was repeated in the ordinary reaction buffer and in the reaction buffer containing detergent. If the percentage inhibition differed markedly between the detergent and ordinary conditions, then this was a suggestion that a compound was a non-selective or promiscuous inhibitor.

Where time and resources permitted, several of the confirmed hits were procured for further characterisation. Another part of the hit validation effort is to determine a specific mechanism of action, so MOI studies were also carried out on some small molecule inhibitors. The principal usefulness of fragments is in uncovering their binding sites, so detailed MOI studies were not conducted on them.

Once greater quantities of the compounds were available, the IC_{50}s of the available hits were determined again. For fragments, they were determined at a range of concentrations between 10 and 500µM, depending on the prior potency data. Each experimental well was matched with a control well, containing an equivalent volume of DMSO. The assay was otherwise carried out using the same conditions as in the screening assay. IC_{50}s for small molecules were determined at saturating concentrations of substrates, as their generally higher potency meant that less material was acquired for each assay. Several of the small molecules showed evidence of time-dependent inhibition, so fixed concentrations of the relevant inhibitors were incubated for increasing amounts of time, before the reaction was initiated and measured using a FLUOstar Omega plate reader. The incubation time required for these inhibitors informed the judgement that a pre-incubation time of 2 hours would be sufficient to ensure that all time-dependent inhibitors would exhibit their full effect in any assay.

Due to the limited supply of some of the small molecules, the MOI assays described above were modified slightly. A mixture of 10µl 7.9 µM (0.3mg/ml) TDH, reaction buffer (containing 1mM KCl) and a range of concentrations of one of the substrates was made in each well. The L-threonine concentrations tested were 1mM, 5mM, 10mM, 12.5mM, 15mM 20mM 22.5mM, 25mM and 30mM. The NAD^+ concentrations tested were 0.1mM, 0.5mM, 1mM, 2mM, 3mM, 5mM, 7.5mM and 10mM. All conditions were compared with a control experiment, containing an equivalent volume of DMSO. The reaction was mixed by five cycles of aspiration and dispensing using a micro-pipette. The reaction was then initiated by addition of 10µl of 100mM NAD^+, or 30µl of 100mM L-threonine solution. The reaction was monitored over 22 seconds, as in the screening assays. To preserve the available compounds, only four inhibitor concentrations were used. The data were fit to MOI equations as described earlier, using Prism (Version 6, 2014, GraphPad Software, http://www.graphpad.com). MOI studies of time-dependent inhibitors were carried out using the same format, rather than by establishing k_{obs}. For those compounds, they were incubated for 1.5 hours at 2-8°C to prevent loss of activity, and then at room temperature for 30 minutes, prior to starting the assay. The data for time-dependent inhibitors were analysed in the same way as for non-time-dependent inhibitors.

2.4.3 Analysis

Two-dimensional chemical/molecular fingerprints are some of the most popular molecular descriptors used in computational chemistry(247–249). Chemical fingerprints are representations of a molecule by use of a bit-string and can be constructed in a variety of ways. The most common form of chemical fingerprint is a path-based fingerprint, where the presence

of atom and bond types along a linear path is denoted by a particular number. The program generatemd in the JChem package(250) was used to generate chemical fingerprints for all unique molecules screened. The construction of chemical hashed fingerprints is shown in Figure 2.4.2 below.

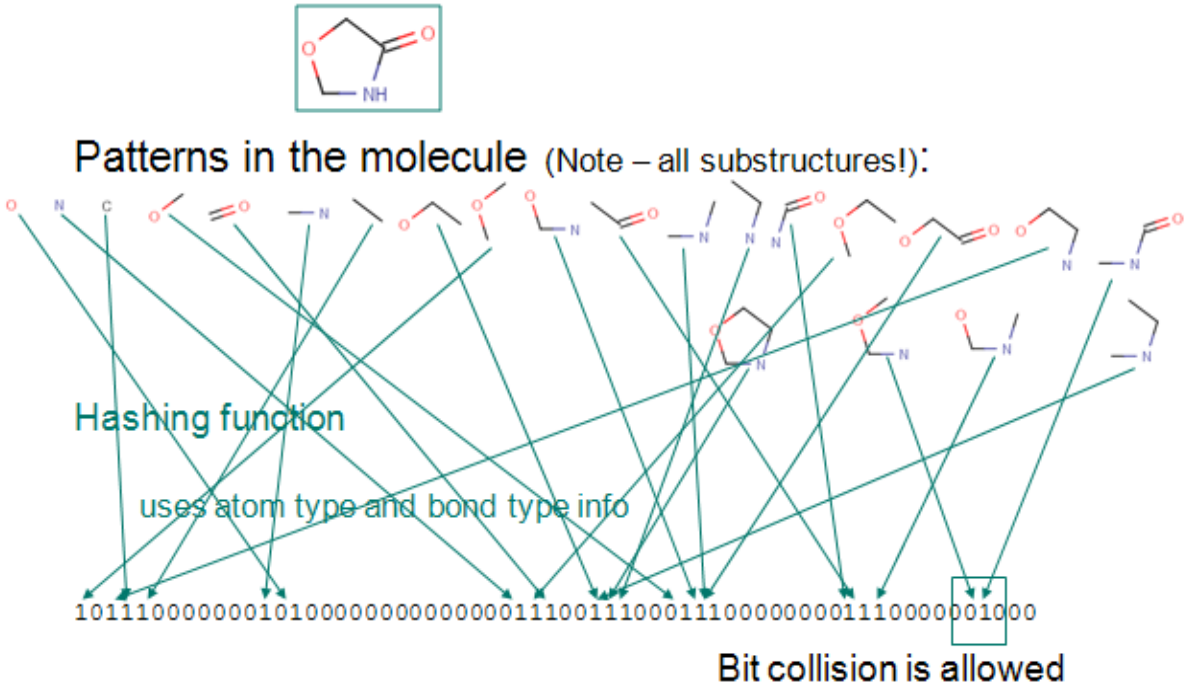


Figure 2.4.2 - the construction of two-dimension path-based binary fingerprints using JChem. The presence of different atom types at certain positions in the model is represented by 1s or 0s in the binary code. Patterns of atoms along linear paths, as well as the bond types that join them, are also represented in the code. The figure shows how the presence of various atoms, bonds and patterns are represented by 1s in the binary code.

Fingerprints were produced in binary format, as described in Figure 2.4.2.The path length was set to 7, the number of bits representing a pattern was set to two and the number of bits in the fingerprint was set to 512, supported by observations of the programmers(251). The Tanimoto coefficient, a metric measuring similarity between two compounds(247,249,252), was generated by the program screenmd, also in the JChem package. The widely applied Tanimoto coefficient is calculated using an equation of the form:

$$T_{(A,B)} = \frac{c}{a + b - c} \qquad \textbf{21}$$

a is the number of "on" bits ("1) in the fingerprint of molecule *A*, *b* is the number of "on" bits in molecule *B* and *c* is the sum of "on" bits in both molecules. *T* is a number between 0 and 1 representing similarity between two molecules, with 1 meaning total similarity(247,248). The program screenmd converts this to a dissimilarity score(253).

$$T_{dissim(A,B)} = 1 - T_{sim(A,B)} \qquad \textbf{22}$$

Higher numbers represent more dissimilar molecules. The screening library and its component libraries were tested for internal dissimilarity by performing an all vs. all comparison carried out

using screenmd. Values were averaged to determine average dissimilarity. This analysis was repeated using pharmacophore fingerprints, also generated in generatemd.

Although the Tanimoto coefficient is widely used, it has the drawback of reducing similarity measures of complex compounds to a single number. To obtain a view of how the molecules are distributed across chemical space, the dissimilarity between the fingerprints was visualised using the Data Visualisation Modelling System (DVMS)(254). This program uses statistical models to produce two-dimensional representations of sets of data. The NeuroScale model(255), uses a Radial Basis Function (RBF)neural network, where the distribution of data points is determined by distances from a number of RBF 'centres'. The equation describing the error function of the network is shown in Figure 2.4.3 below.

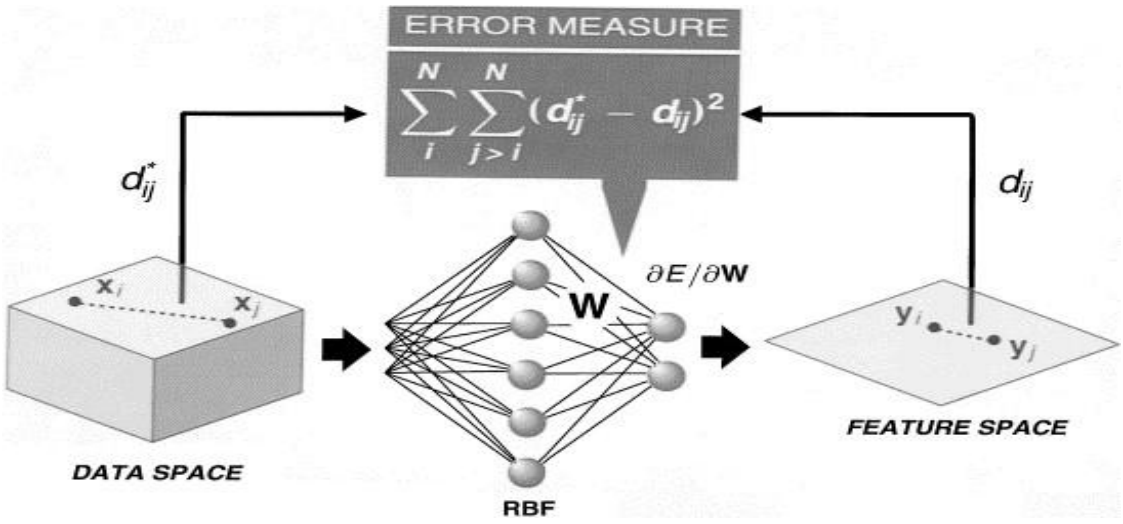


Figure 2.4.3 - the NeuroScale model and it's use of the Radial Basis Function to define distances between data.

N is the number of molecules in the data set, d_{ij}* is the distance between molecules *i* and *j* in the data, and d_{ij} is the distance between them in the visualisation space. d_{ij} is given by the Soergel distance function, which is equivalent to T_{dissim} above(254). A subset of the compounds tested in virtual screening is shown in a data visualisation below, as an example. Note that the relative distances between each data point depends on the calculated differences between the molecules that they represent. Also note that there are no axes on such visualisations.

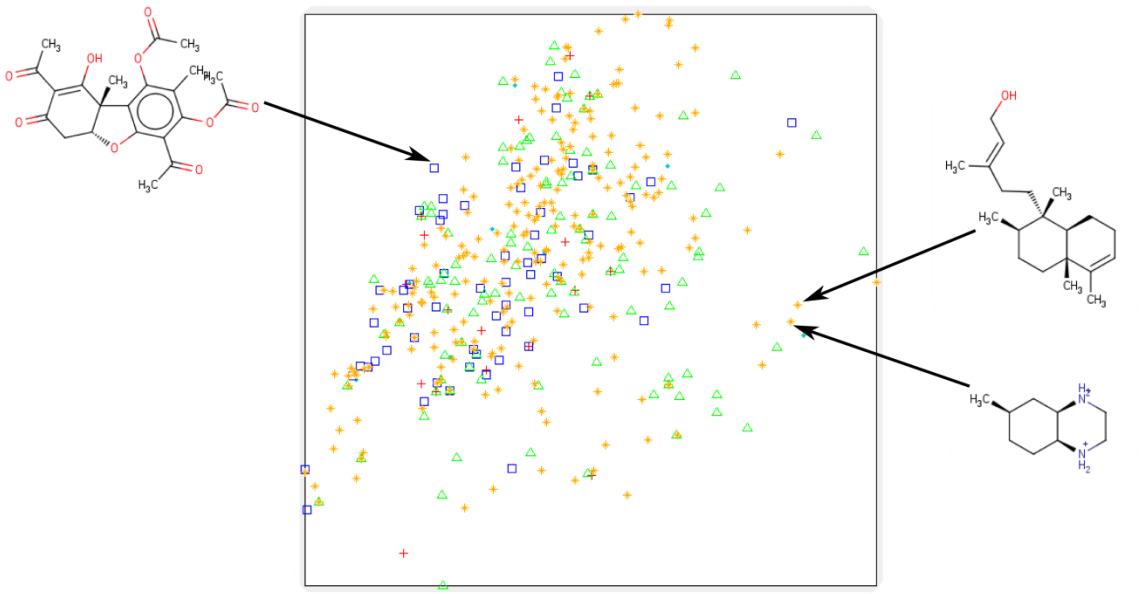


Figure 2.4.4 - a visualisation of a subset of the virtual screening library produced by the DVMS. Each data point represents the chemical fingerprint of one molecule. The relative positions of all data points are determined by the differences between the molecules that they represent. To illustrate this, a few representative molecules are shown with arrows pointing to the data points that they represent.

2.4.4 Trypanosome growth inhibition

Confirmed TDH inhibitors were supplied in stock solutions of 10µg.ml^{-1} for testing against bloodstream form trypanosomes (*T. b. brucei* strain 427, variant 221) in culture. These assays were carried out by Professor Jon Kelly at the London School of Hygiene and Tropical Medicine. Activity was reported as percentage inhibition of trypanosome growth.

2.5 Commercialisation of NTD drugs

The purpose of this chapter is to detail the various ways in which drugs and vaccines for neglected tropical diseases (NTD) are developed and advanced towards the market. This chapter will explore factors affecting product (drug/vaccine) development for NTDs and will indicate how they apply to the development of a drug to treat HAT or Chagas Disease. There are a few excellent resources that inform on the funding available for NTD product development. One such resource is the Global Funding of Innovation for Neglected Diseases (G-FINDER) report(66) and database(73), the results of a survey on funding for NTD research & development (R&D), compiled by Policy Cures. Another useful resource is BIO Ventures for Global Health's (BVGH) Global Health Primer(74), a database that holds details on NTD products currently in development and the organisations involved in those development projects. In addition to these, the research underlying this chapter was guided by a number of reviews of NTD product development activity(50,52,57,60). However, there appears to be a

gap in the research, mainly pertaining to the different strategies used by researchers and developers to advance products towards the market and the availability of funding to advance development at specific stages in the pharmaceutical research and development (R&D) lifecycle. Furthermore, the literature suggests that the formation of partnerships between different organisations is almost ubiquitous amongst NTD product development campaigns, although there is a scarcity of reviews providing specific details on such partnerships. Therefore, the research presented will cover the following areas: the different commercialisation strategies for new products to control NTDs, the level of evidence required when entering into partnerships, and the nature of partnerships formed during pharmaceutical development for NTD products. The implications for NTD product developers were considered from the point of view of researchers in academic and public or not-for-profit research institutions (PSNRI), and recommendations on the most appropriate strategies to be utilised in the pursuit of the research discussed in this thesis will be given.

In order to identify the gaps in the literature regarding NTDs, a literature search was conducted using combinations of the following keywords: “neglected tropical disease”, “NTD”, “drug” and “development”. The basis of much of the background information supporting the research came from reviews of different aspects of NTD product development. This background information was supplemented with specific examples of successful development projects in this area. In order to undertake a more complete exploration of the topic, it was also important to gather more data relating to pharmaceutical discovery and development in general, as much of this is transferable to the development of NTD drugs. In particular, information was sought on the originators of marketed drugs, and on the various parties involved in their development. As the development of drugs for NTDs remains a special case within the pharmaceutical industry, examples of development strategies in other industries, where similar barriers to development are present (e.g. a lack of market incentive), were sought. This involved gathering information from sources outside of the pharmaceutical and biomedical literature.

3. Results

3.1 Structural Information

3.1.1 Protein Expression and Purification

Both BL21 (DE3) and Rosetta™ competent cells were successfully transformed with TDH and KBL DNA using the methods described in Section 2.1.1. Expression tests also showed that selective expression of the relevant genes was inducible by the addition of isopropyl ß-D-thiogalactopyranoside (IPTG). Figure 3.1.1 shows an SDS-PAGE gel that analyses an expression test for TDH-expressing bacteria.

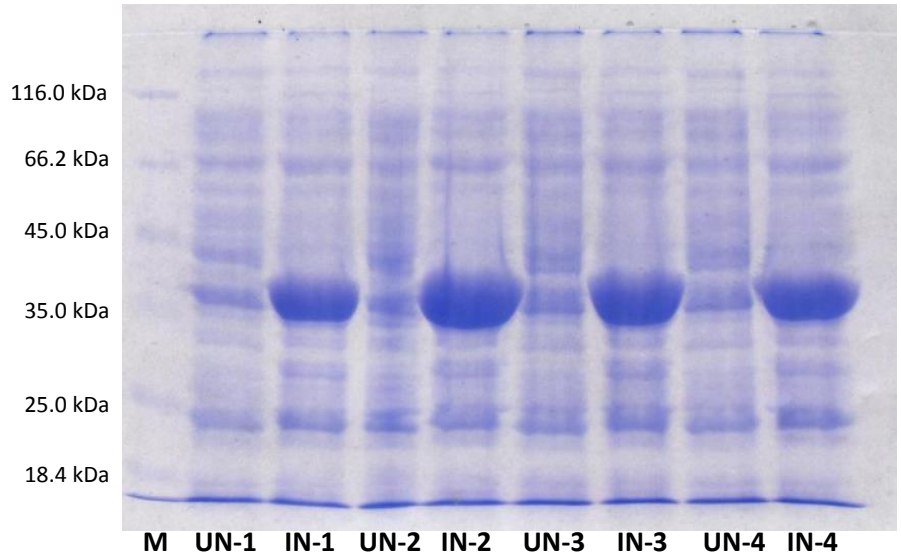


Figure 3.1.1 - SDS PAGE of a successful expression test for TDH. M = standard molecular weight marker, UN = un-induced cell and IN = induced cell. Four colonies, numbered 1 to 4, were tested and the presence of TDH was confirmed by a large protein band corresponding to a protein of around 38 kDa in mass. This is an example of a succesful expression test, as the cells were induced to over-express TDH, whilst the expression of other proteins remained constant.

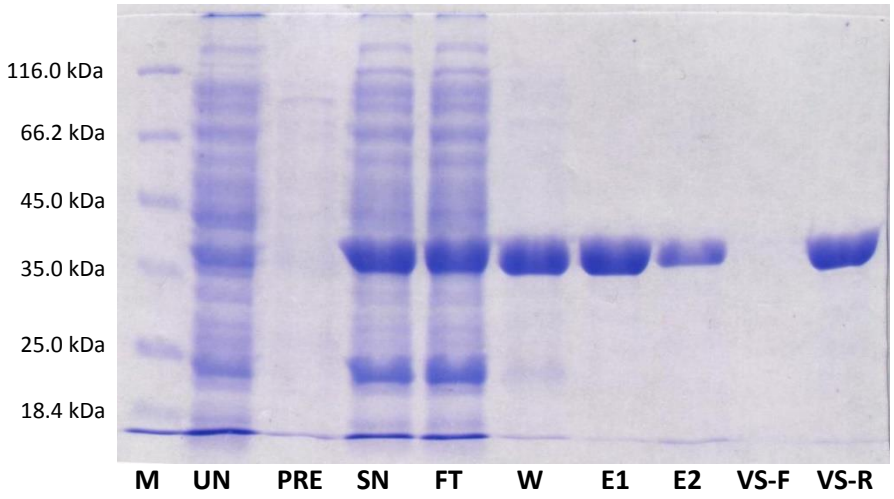


Figure 3.1.2 - SDS-PAGE analysis of the purification of TDH from the supernatant obtained through protein extraction procedures. M = standard molecular weight marker; PRE = solution eluted from Ni-NTA column prior to protein loading; SN = supernatant; FT = 'flow-through', solution flowing through the Ni-NTA column during initial loading of the supernatant; W = 'wash', the solution eluted from the column upon washing with binding or wash buffer; E1 = eluate collected following elution with 5ml of elution buffer; E2 = eluate following an additional 5ml of elution buffer; VS-F = 'Vivaspin filtrate', formed during ultrafiltration in a Vivaspin™ tube to concentrate TDH; VS-R = 'Vivaspin retentate', the solution retained during ultrafiltration to concentrate TDH. This analysis shows that some protein was lost in the flow-through and wash, suggesting that the amount of protein exceeded the binding capacity of the Ni-NTA column. Looking at lanes E1 and E2 also reveals that 5ml is actually insufficient to elute all of the protein from column, contrary to the manufacturer's recommendations. This may have been due to the fact that the quantity of TDH was above the column's binding capacity. Overall, this purification procedure was very successful, as TDH was purified from the supernatant, which contained a multitude of other proteins.

When TDH was extracted from induced, cultured bacteria, a high concentration of soluble TDH was found in the supernatant after extraction, although a significant amount of protein was present in the insoluble cell pellet (see lane 'CP' in Figure 3.1.3). Purification of TDH from the supernatant by affinity chromatography was very effective and was able to isolate highly pure protein (see Figure 3.1.2) in most cases. However, an analysis of the SDS-PAGE result in Figure 3.1.2 shows that TDH was found in the solution that flowed through and was washed through the nickel column when following the procedure for affinity chromatography that is described in the methods. This indicates that the quantity of TDH present in the supernatant exceeded the capacity of the column. For this reason, the procedure was modified and two columns were attached end to end. This ensured that protein was not lost as supernatant was initially loaded onto the column (see Figure 3.1.3). However, some TDH was still lost in the wash step. It was necessary to repeat the process, treating the wash fraction as one would the supernatant, in order to recover that protein.

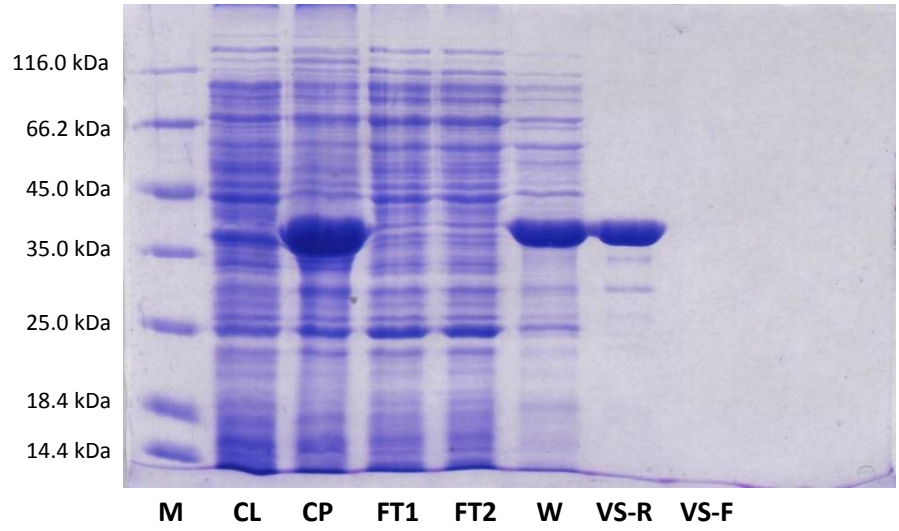


Figure 3.1.3 - SDS PAGE analysis of a modified purification procedure. M = standard molecular weight marker; CL = un-induced cell lysate; CP = cell pellet, containing insoluble protein; FT1/2 = flow through, solution flowing through the two connected Ni-NTA columns; W = wash, solution eluted from the column following washing with binding or wash buffer; VS-R = Vivaspin retentate; VS-F = Vivaspin filtrate. Although a large amount of the expressed TDH was present in the insoluble cell pellet, a reasonable amount of protein was isolated from the supernatant (not analysed here). Connecting two Ni-NTA columns together meant that all of the protein was initially retained on the columns. However, some of the protein was still washed away with binding buffer. Nevertheless, a large amount of TDH was ultimately eluted from the column and concentrated by ultrafiltration.

The protein extraction and purification processes were also effective for the purification of KBL, as shown in the figures below. Figure 3.1.4 shows an expression test for KBL and Figure 3.1.5 analyses the extraction of soluble KBL from KBL-expressing bacterial cell cultures. Finally, Figure 3.1.6 confirms the purification of KBL by affinity chromatography.

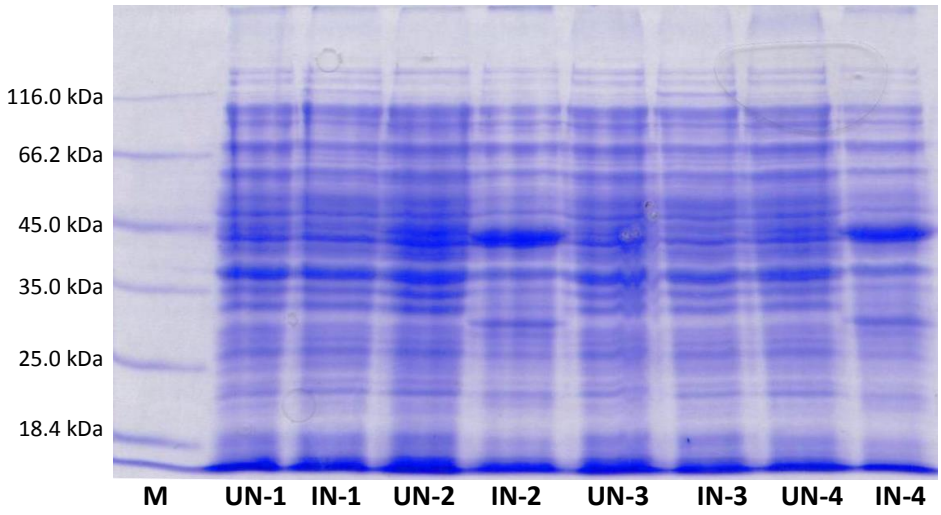


Figure 3.1.4 - SDS PAGE of a successful expression test for KBL. M = standard molecular weight marker, UN = un-induced cell and IN = induced cell. Four colonies, numbered 1 to 4, were tested and the presence of a large protein band corresponding to a protein of around 46 kDa in mass was used to confirm successful induction. The over-expression of KBL was only inducible in colonies 2 and 4, in this example.

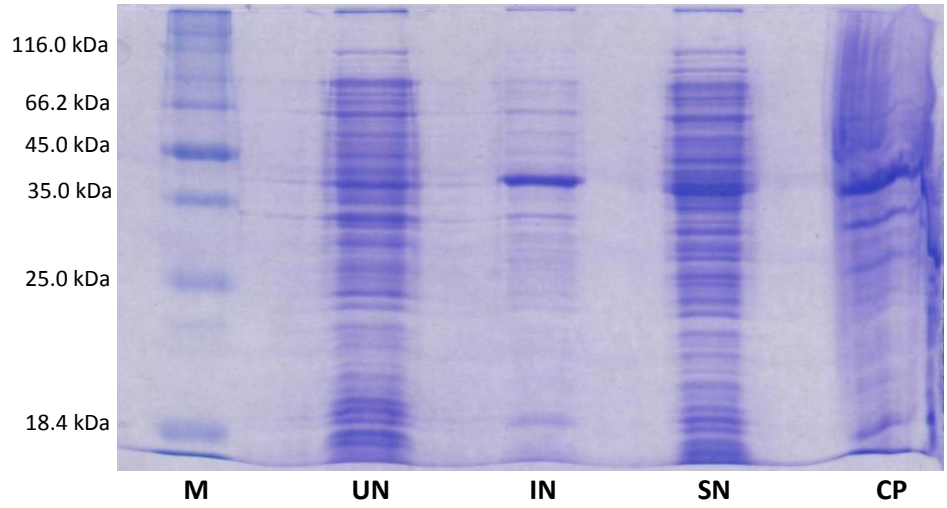


Figure 3.1.5 - Extraction of KBL from an induced cell culture. UN = un-induced cell; IN = induced cell; SN = supernatant; CP = cell pellet. The presence of KBL was confirmed by the presence of a protein band corresponding to a molecular weight of approximately 46 kDa. The analysis above shows that KBL was selectively induced and that it was present in both the supernatant and the cell pellet.

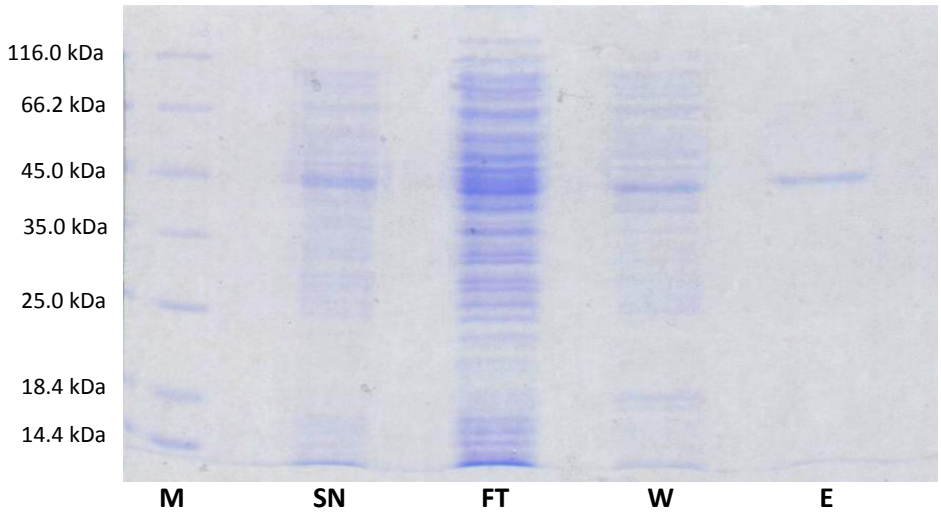


Figure 3.1.6 - SDS PAGE analysis of KBL purification by affinity chromatography. M = standard molecular weight marker; SN = supernatant; FT = flow-through, the solution that flowed through the Ni-NTA column upon loading of the protein; W = wash, the solution obtained after washing the column with binding or wash buffer; E = eluate, the solution eluted from the column by applying elution buffer. This example shows the successful isolation of a pure KBL sample from the supernatant.

In Figure 3.1.3 and Figure 3.1.7, it can be seen that there were some impurities in the final protein sample. This occurred in a minority of cases, but when it did happen the most common impurity had a molecular weight of around 66-70 kDa. This matches the molecular weight of the *E.* coli protein known as glucosamine-6-phosphate synthase or GlmS, which is reported to have a molecular weight of 67 kDa. It is highly likely that GlmS was the contaminant, as it is known to co-purify with His-tagged proteins during affinity chromatography carried out using nickel columns(256). Another common contaminant was sometimes apparent on SDS-PAGE gels, with a protein band corresponding to 25 kDa. The identity of this contaminant is likely to be

Peptidoylproline *cis–trans* isomerase or SlyD, which is another well-known native *E. coli* protein that co-purifies with His-tagged proteins during affinity chromatography(256,257).

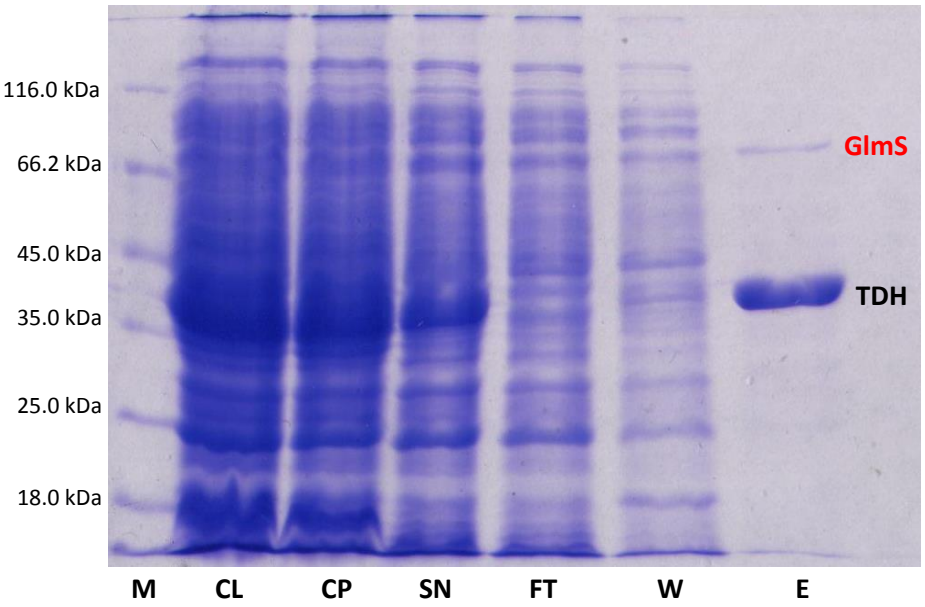


Figure 3.1.7 - SDS PAGE analysis of TDH extraction and purification. M = standard molecular weight marker; CL = induced cell lysate; CP = cell pellet; SN = supernatant, FT = flow-through, the solution that flowed through the Ni-NTA column when loading the protein; W = wash, the solution washed from the Ni-NTA column with binding or wash buffer; E = eluate, the solution obtained following application of elution buffer to the column. In this example, TDH was present in large amounts in both the cell pellet and the supernatant. The soluble TDH was purified from the supernatant without significant losses of protein in the flow-through or wash fractions (two Ni-NTA columns, attached end-to-end) were used. A small amount of a co-purifying contaminant, GlmS, was present in the final sample.

For the purpose of removing impurities such as GlmS and SlyD, and to ensure a purer protein sample, ion exchange chromatography proved to be useful. This is evident in the SDS-PAGE analysis in Figure 3.1.8. When purifying TDH by this method it was observed that some of the protein was lost in the unbound fraction that flowed through the column shortly after column loading. This suggests that the conditions of the equilibration or binding buffer were not optimal for the initial loading of the protein onto the column. Despite this, the method did prove to be effective for purifying the protein of interest.

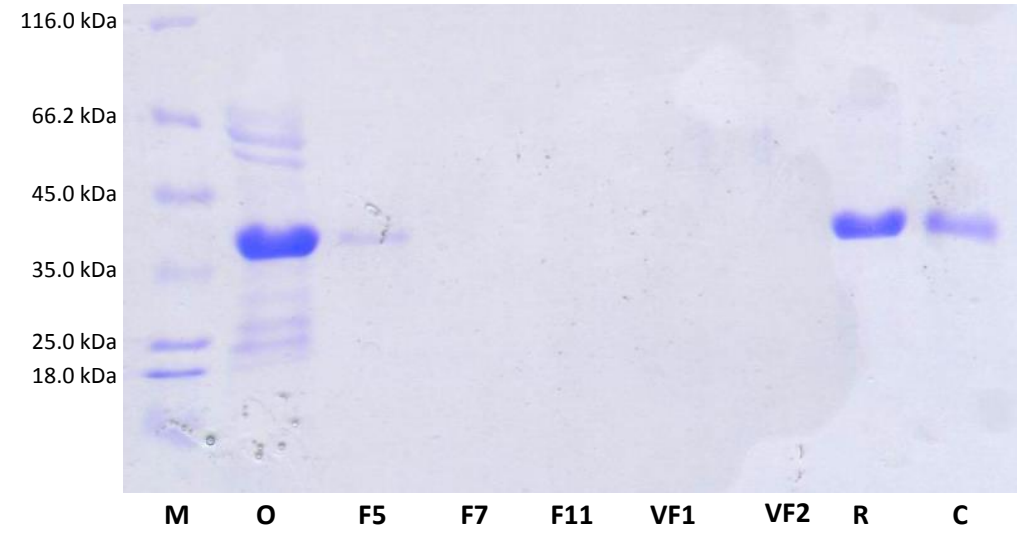


Figure 3.1.8 - SDS-PAGE analysis of TDH purification by ion exchange chromatography. M = standard molecular weight marker; O = original impure TDH sample; F5 = fraction 5, containing solution corresponding to 50-60ml elution volume and including some of the un-bound protein fraction; F7 = fraction 7, containing solution corresponding to 70-80ml elution volume; F11 = fraction 11, containing solution corresponding to 110-120ml elution volume; VF1/2 = Vivaspin filtrate, the filtrate obtained following ultrafiltration of the purified TDH sample; R = Vivaspin retentate; C = control, a previously purified TDH sample. Although the ion exchange chromatography procedure was useful, the presence of TDH in the early fractions, which corresponded to the tail end of a peak representing the un-bound fraction, suggests that a significant amount of protein did not bind to the column.

Expression and purification of TDH using the methods described was able to produce expression yields of up to 1.48% (mass of protein following affinity chromatography, as a percentage of wet bacterial cell mass). Similar quantities of purified KBL were obtained using the same processes, although no yield was calculated. The type of competent cell (BL21 or Rossetta™ cells) used did not appear to affect expression yields significantly. The use of recently transformed bacteria, rather than cells that had been stored for long periods of time (more than one month), also seemed to correlate with purer final protein samples.

The protein samples appeared to be stable when stored for long periods of time in storage buffer at approximately 4°C. After storage periods of more than one month, small amounts of white precipitate could be seen at the bottom of storage vessels. However, after filtering the protein solutions, the UV absorbance spectra appeared the same, and there was no significant loss of activity (for TDH). It was possible to store the protein solutions for several months by freezing them at -20°C. Upon thawing, the UV absorbance spectra remained the same, and there was no significant loss of enzyme activity observed during assays of TDH.

3.1.2 Crystal growth and data collection

Diffracting TDH crystals were successfully grown under a number of different conditions. Most of the crystallisation solutions in which TDH was grown contained polyethylene glycol (PEG) 4000 (4K) as the precipitant. The conditions in which crystals that were used to obtain X-ray data were grown are listed in Table 3.1.1, along with data collection and refinement statistics for the related data sets. No diffracting KBL crystals were produced by the conditions tested.

Crystals that grew in some of the conditions tested were confirmed as salt crystals by observing the X-ray diffraction pattern.

The original protocol employed for cyro-cooling of crystals involved cryoprotection with 20-25% v/v glycerol (see section 2.1.3). A number of TDH crystals were cryo-cooled, replacing glycerol with PEG 400 or ethylene glycol, or using no cryoprotectant at all, to determine whether this would affect the quality of data that could be collected from such crystals. It was found that quality data could be collected from crystals that were cooled with or without cryoprotectant. Subsequently, the X-ray models labelled TNADH4 and TM131 were derived from data collected from crystals that were not cryoprotected prior to cryocooling. Quality data was also collected from a crystal that was cryoprotected with 20-25% PEG 400. Data related to that crystal did not provide any new insights relevant to this study, so it is not presented in this thesis.

Table 3.1.1 - details of the crystallisation, data collection and structure solution of TDH. Each crystal and data set was assigned an identity code, which is indicated in the second header row of the table.

	Model						
	TK8	**TS2**	**TIO8**	**TQ5**	**AAT1**	**TNADH4**	**TM131**
Crystallisation Conditions							
Conditions	- 0.2M lithium sulphate, 0.1M Tris, 30% w/v PEG 4K, pH 8.5 - TDH (6.4 x 10^{-2}mM [2.4mg/ml]), KBL (3.8 x 10^{-2}mM[1.8mg/ml]), NAD^+ (0.09mM), PLP (0.18mM), L-Threonine (7.33mM)	- 0.1M MES, 0.2M ammonium sulphate, 30% w/v PEG 5K MME, pH 6.5 - TDH (9.0 x 10^{-2}mM [3.4mg/ml]), L-serine (50mM)	- 0.1M tri-sodium citrate, 30% w/v PEG 4K, 0.2M ammonium acetate, pH 8.5 - TDH (3.7 x 10^{-2}mM [1.4mg/ml]), NAD^+ (10mM), BPOB (0.5mM)	- 0.1M tri-sodium citrate, 30% w/v PEG 4K, 0.2M ammonium acetate, pH 5.6 - TDH (3.7 x 10^{-2}mM [1.4mg/ml]), NAD^+ (8.5mM), quinine (2.5mM [hydrochloride dehydrate])	- 0.1M tri-sodium citrate, 30% w/v PEG 4K, 0.2M ammonium acetate, pH 5.6 - TDH (10.4 x 10^{-2}mM [3.9mg/ml]), NAD^+ (10mM), L-*allo*-threonine (30mM)	- 0.2M sodium acetate, 0.1M Tris, 30% w/v PEG 4K, pH 8.5 - TDH (9.3 x 10^{-2}mM [3.5mg/ml]), NADH (10mM), L-Threonine (30mM)	- 0.1M tri-sodium citrate, 25% w/v PEG 4K, 0.2M ammonium acetate, pH 5.6 - TDH (3.7 x 10^{-2}mM [1.4mg/ml]), NAD^+ (10mM), methylglyoxal (8mM)
Cryoprotectant	Glycerol	Glycerol	Glycerol	Glycerol	Glycerol	None	None
Data Collection							
Beamline	ESRF	DLS I04-1	DLS I04-1	DLS	DLS I04-1	DLS I02	DLS I04-1
Wavelength (Å)	0.9334	0.9173	0.9173	0.9173	0.92	0.9795	0.92001
Space group	$P2_12_12$	$P4_32_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$

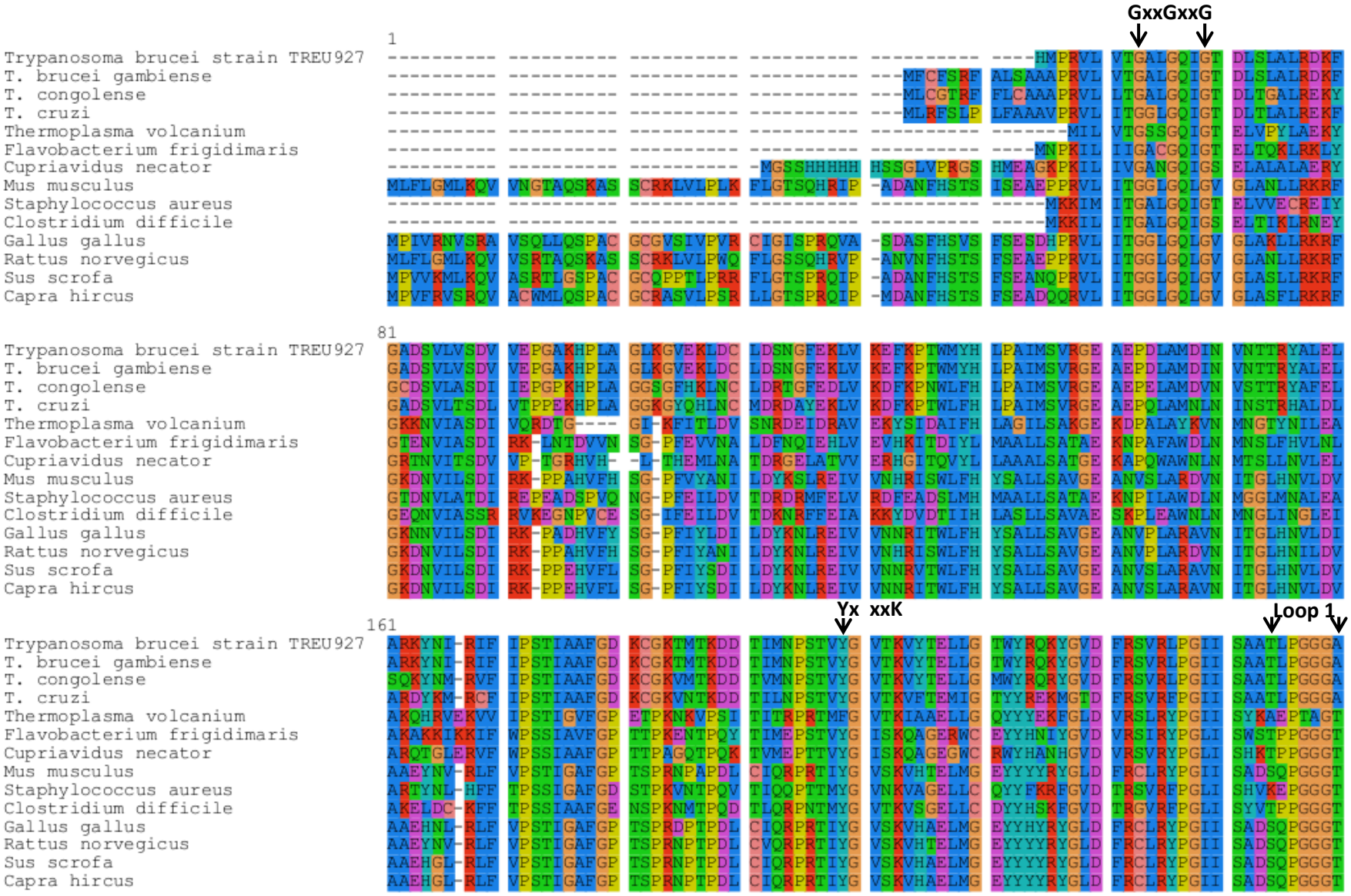
	Model						
	TK8	**TS2**	**TIO8**	**TQ5**	**AAT1**	**TNADH4**	**TM131**
Unit cell parameters							
a (Å)	132.04	91.76	90.40	90.07	133.03	83.45	133.45
b (Å)	276.49	91.76	131.53	133.10	273.06	136.12	278.63
c (Å)	55.74	93.60	55.02	55.61	55.80	55.69	56.27
α (°)	90	90	90	90	90	90	90
β (°)	90	90	90	90	90	90	90
γ (°)	90	90	90	90	90	90	90
Max. resolution (Å)	2.3	2.1	2.2	1.9	2.26	1.72	1.77
No. of reflections	624537	301645	109277	203173	297998	435135	438691
No. of unique reflections (used)	93173	23982	32475	61881	91289	67909	193067
Solvent content (%)	46.34	52.83	42.88	43.94	44.98	41.24	46.7
Matthews coefficient ($Å^3.Da^{-1}$)	2.29	2.61	2.15	2.19	2.23	2.09	2.31
Resolution range (Å)	48.143 - 2.3	29.073 - 2.1	28.226 - 2.2	28.552 - 1.9	31.09 - 2.26	46.32-1.72	39.60 - 1.77
Multiplicity	6.7	12.6	3.4	3.3	1.9	6.4	6.2
Completeness (%)	99.188	99.98	95	97.7	90.4	99.5	99.6
Mean I/σ	11.2	33.6	13.4	12.7	5.6	11	12.6
R_{merge}	0.12	0.054	0.067	0.074	0.151	0.094	0.121
R_{pim}	0.049	0.016	0.04	0.048	0.181	0.04	0.055
R_{meas}	0.13	0.056	0.079	0.088	0.214	0.102	0.133
Refinement							
R_{work}	0.1746	0.1741	0.1643	0.1864	0.2073	0.1522	0.16071
R_{free}	0.2618	0.2400	0.2411	0.2321	0.2809	0.1976	0.20074
RMS bond lengths (Å)	0.0178	0.0245	0.0201	0.0241	0.0128	0.0198	0.0201
RMS bond angles (°)	1.764	2.059	1.826	1.99	1.6492	2.1037	1.9749
Average B-factor ($Å^2$)	21.09	26.769	26.234	22.138	33.76	27.689	19.218

	Model						
	TK8	**TS2**	**TIO8**	**TQ5**	**AAT1**	**TNADH4**	**TM131**
Ramachandran favoured, % of total	1830, 96.62%	309, 96.87%	613, 96.69%	618, 97.48%	1801, 95.14%	622, 97.65%	1824, 97.91%
Ramachandran allowed, % of total	61, 3.22%	8, 2.51%	18, 2.84%	14, 2.21%	82, 4.33%	13, 2.04%	33, 1.77%
Ramachandran outliers, % of total	3, 0.16%	2, 0.63%	3, 0.47%	2, 0.32%	10, 0.53%	2, 0.31%	6, 0.32%
Molecules in the asymmetric unit	6	1	2	2	6	2	6
Quaternary structure	Dimer	Dimer	Dimer	Dimer	Dimer	Dimer	Dimer
Residue range	Pro3-Leu321	His1-Leu321	Pro3-Leu321	Pro3-Leu321	Pro3-Leu321	His1-Leu321	Met2-Leu321
No. protein atoms per monomer	2493	2532	2493	2493	2500.5	2507	2533
No. water molecules per monomer	252.3	152	220.5	100.5	73.7	210	241.1
NAD^+ bound?	Yes	No	Yes	Yes	Yes	Yes (NADH)	Yes
Ligands bound, Mean B-factor	-	-	-	-	L-*allo*-threonine, 39.53 L-*allo*-threonine, 26.88	L-threonine, 26.06 L-threonine, 30.58	-
Other solvent molecules/ions	glycerol, sulphate	glycerol	glycerol, acetate, Na^+	glycerol, acetate	glycerol, acetate	glycerol, Na^+	glycerol, acetate

The TDH structure models obtained through X-ray crystallography provide many insights on the structure and function of the enzyme. As it was possible to co-crystallise TDH with NAD and other substrates, information on the binding of TDH with its substrates has also been gained. The structure of TDH and how this may relate to its function is discussed below, in light of the crystallographic data, as well as data from other experimental techniques.

3.1.3 Primary Structure

The protein and nucleotide BLAST searches (see Section 2.1.3.3) of TDH amino acid sequences and nucleotide sequences, respectively, presented several similar TDH proteins from other species that had high sequence similarities to TDH from *Trypanosoma brucei*. These proteins typically share approximately 40% sequence similarity with TDH from *T. brucei*. The sequences included some that correspond to previously studied TDH enzymes from other species. The published amino acid sequences of these proteins are aligned with that of TDH from *T. brucei* TREU 927 strain in Figure 3.1.9. Sequences of TDH from other *Trypanosoma* species or subspecies are also included.


GxxGxxG
Yx xxK
Loop 1
Trypanosoma brucei strain TREU927
T. brucei gambiense
T. congolense
T. cruzi
Thermoplasma volcanium
Flavobacterium frigidimaris
Cupriavidus necator
Mus musculus
Staphylococcus aureus
Clostridium difficile
Gallus gallus
Rattus norvegicus
Sus scrofa
Capra hircus

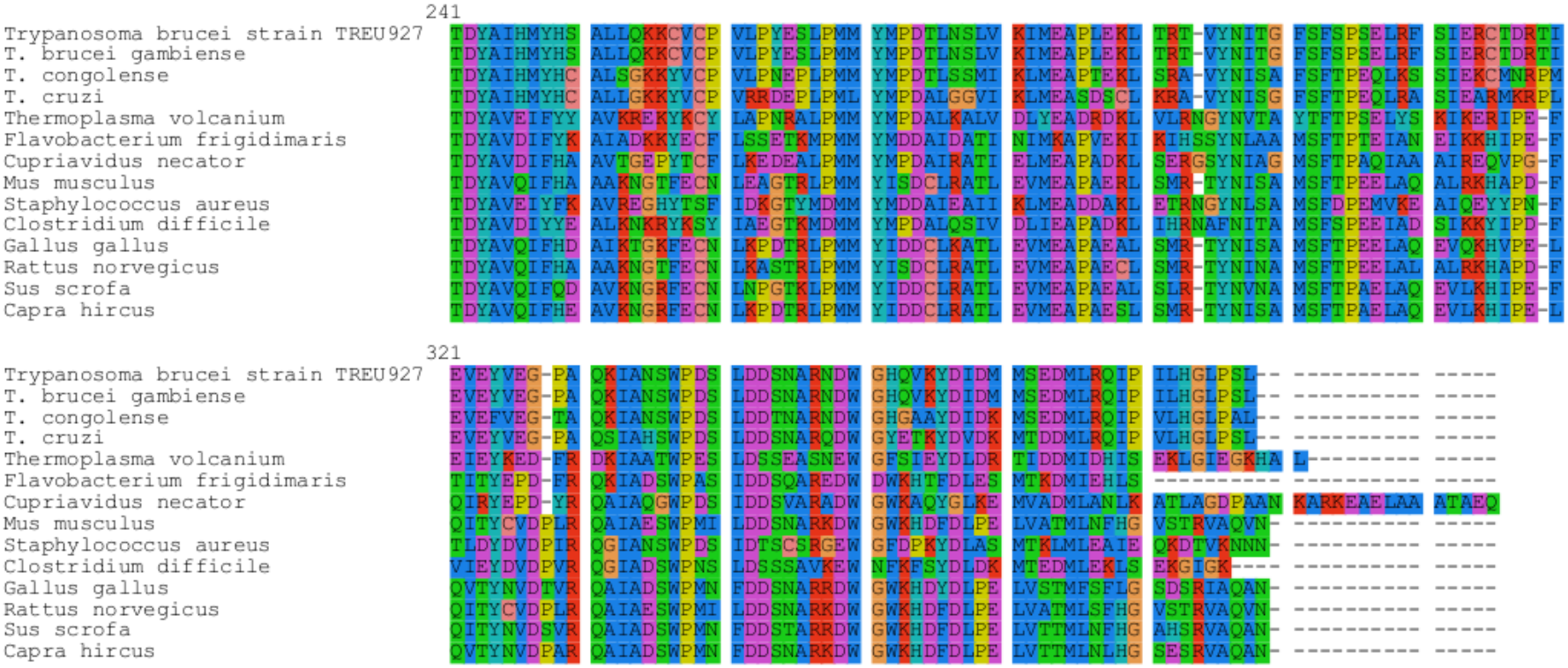


Figure 3.1.9 - amino acid sequence alignments of TDH from different organisms. The sequence of the TDH variant studied in this thesis is listed at the top (*T. brucei* strain TREU 927). Each residue is represented by the corresponding single-letter code for that amino acid. Sequences of non-aligning residues are represented by a dash. Key conserved sequence segments are indicated by an arrow and bold text. The glycine-rich sequence GxxGxxG, and the catalytic motif YxxxK have been highlighted, including a sequence that corresponds to a flexible loop (Loop 1). The alignment was created using the software Seaview 4(258). Sequences were identified through the NCBI protein and nucleotide blast searches(181).

The TDH sequences from *T. brucei* strain TREU 927 and *T. brucei gambiense* were essentially identical, confirming that the TDH studied in this work is relevant to a disease-causing pathogen. The animal disease-causing pathogen *T.* congolense has an unnamed protein that shares 78% sequence identity with the TDH studied here. Therefore, there is a high probability that this protein is a TDH. The sequence identity shared between TDH from *T. brucei* strain TREU 927 and the Chagas disease-causing pathogen, *Trypanosoma cruzi*, is 72%. Thus, the TDH enzymes from a range of *Trypanosoma* species are closely related. The phylogenetic tree in Figure 3.1.10 illustrates this and gives an indication as to the relationship between the enzymes in the species included.

Other published TDH enzymes with a high sequence similarity to the enzyme studied in this thesis are the UDP-galactose 4'-epimerase-like (GalE) TDH enzymes from the bacteria *Thermoplasma volcanium, Flavobacterium frigidimaris* and *Cupriavidus necator*. The corresponding enzymes from these organisms all have monomeric molecular weights reported between 35 and 37.2kDa(124–126,197), matching that of TDH from *T. brucei*. The phylogenetic tree in Figure 3.1.10 indicates some variation between the sequences of TDH from the various bacterial species.

Another group of related TDH enzymes originate from mammalian species: Gallus gallus (chicken), Rattus norvegicus (rat), Sus scrofa (boar/pig) and Capra hircus (goat). Again, TDH enzymes from these species have been studied previously. Reported molecular weights corresponding to TDH monomers from these organisms include 36 kDa for chicken (259) and 37 kDa for pig(260).

Looking at the sequences of the various TDH enzymes shows that they belong to the short chain dehydrogenase/reductase (SDR) superfamily of enzymes. These enzymes catalyse oxidative or reductive reactions and share certain highly conserved structural motifs(261). Two structural features that can be identified from the TDH sequence are highlighted in Figure 3.1.9. The first is a glycine-rich region, which is represented by the sequence GxxGxxG, where the letter x can be replaced by any amino acid. The second feature consists of a tyrosine and a lysine residue separated by three other residues: YxxxK(118). As can be seen in Figure 3.1.9, these motifs are highly conserved across the TDHs in different species.

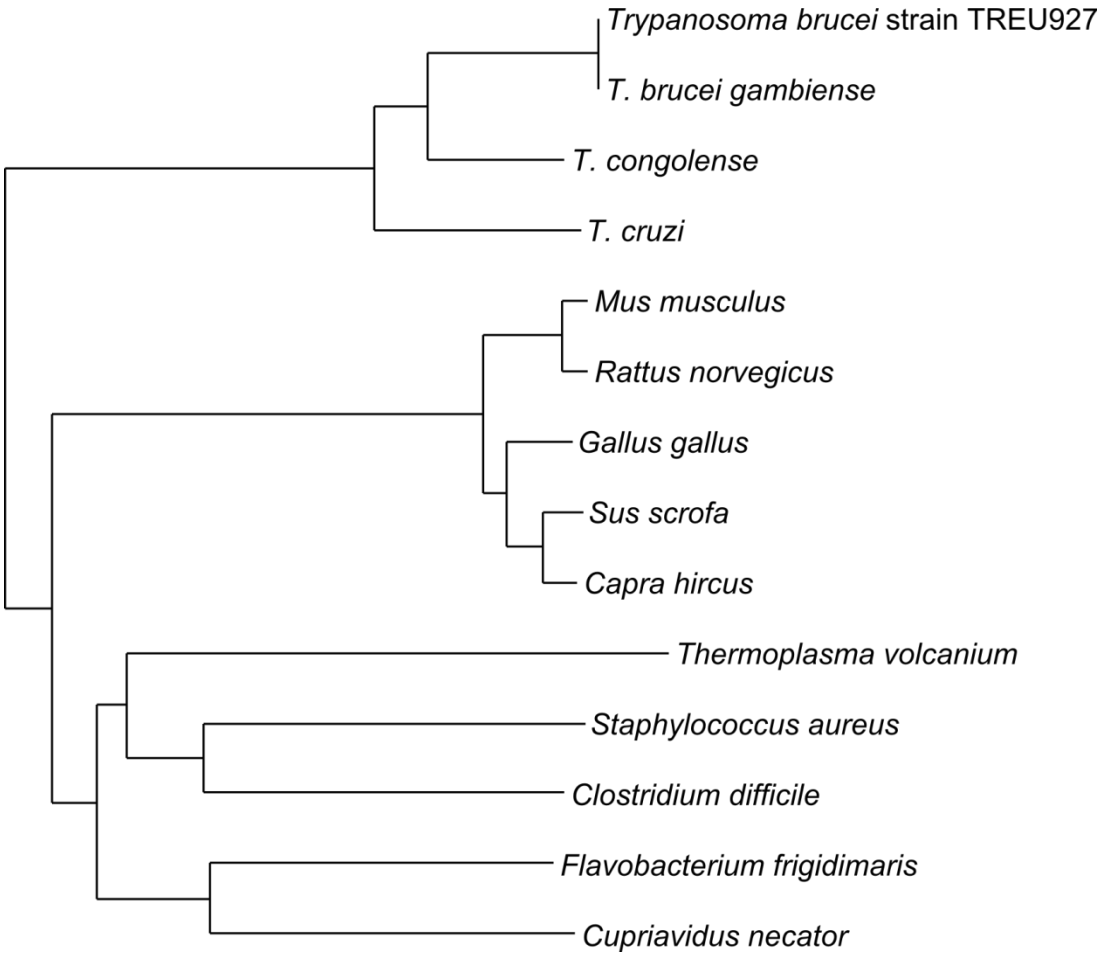


Figure 3.1.10 - phylogenetic tree connecting the different TDH sequences present in different organisms. Different branches linking organisms to the same root indicate that those TDH enzymes have a common genetic ancestry. The lengths of the branches (horizontal) are proportional to the degree of genetic change between species. Hence, *T. brucei* strain TREU927 and *T. brucei gambiense* share identical TDH genes/proteins. The TDH from *Trypanosoma* are grouped together, as are those from mammals and those from bacteria. The genes of bacteria appear to vary to much greater extents than those of *Trypanosoma* and mammals. This phylogenetic tree was created using the default settings of PhyML, which is integrated in Seaview4(258).

The BLAST searches also revealed a number of similar sequences for proteins from other organisms of biological and pharmacological interest. These proteins, and the relevance of the organisms from which they come from are listed in Table 3.1.2

Table 3.1.2 - organisms of pharmacological/medical interest that may possess TDH enzymes similar to that from *T. brucei*.

Sequence Identity	Organism	Role of organism	Confirmed /Putative / Unnamed
46%	*Blastocystis hominis*	Possible role in gastrointestinal disease	Unnamed
45%	*Ascaris suum*	Causes roundworm infection in pigs	Confirmed: TDH
43%	*Phytophthora infestans*	Causes potato blight	Putative: TDH
42%	*Anopheles gambiae*	Malaria vector	Unnamed
43%	*Aedes aegypti*	Dengue, Chikungunya and Yellow fever vector	Putative: TDH
41%	*Culex quinquefasciatus*	Lymphatic filariasis vector	Confirmed: TDH
41%	*Glossina morsitans morsitans*	Tsetse fly	Unnamed
42%	*Enterococcus faecium*	Normal gut flora, but important source of nosocomial infections	Putative: epimerase/reductase
41%	*Pediculus humanis corporis*	Human body louse	Putative: TDH
42%	*Staphylococcal* spp. (incl. *Staph aureus*)	Several human infections	Unnamed
41%	*Mus musculus*	Study with inhibitors of mouse TDH (Alexander et al. 2011). Role in embryonic stem cells.	Confirmed: TDH
28%	*Haemophilus influenzae*	Human diseases, including meningitis	Uncharacterised

3.1.4 Secondary Structure

X-ray models of TDH revealed that the structure is made up of a mixture of α-helices, β-sheets and disordered loops. When observed visually, the distribution of secondary structure elements in TDH appears to be approximately 38% α-helices and 12% β-sheets, with the remaining 50% of the protein not forming specific folds.

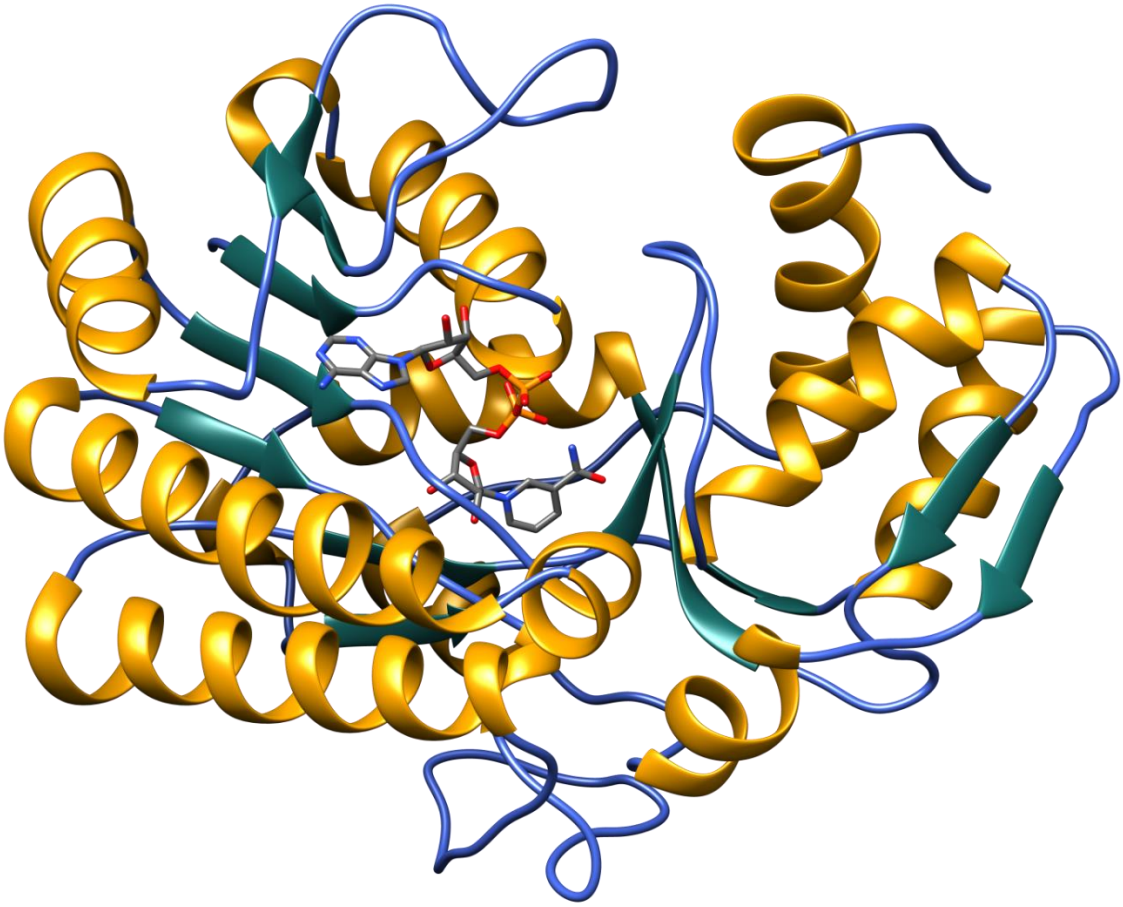


Figure 3.1.11 - Ribbon representation of a TDH monomer coloured by secondary structure. The cofactor NAD is shown in a stick representation.

3.1.5 Tertiary Structure

The TDH monomer can be divided into two domains: an NAD-binding domain, which includes the first 180-200 amino acids, and a catalytic domain at the C-terminal end of the protein. Though, in three dimensions, there is some crossover between the domains and the different portions of the amino acid sequence. The arrangement of secondary structure elements is depicted in the protein topology diagram in Figure 3.1.12.

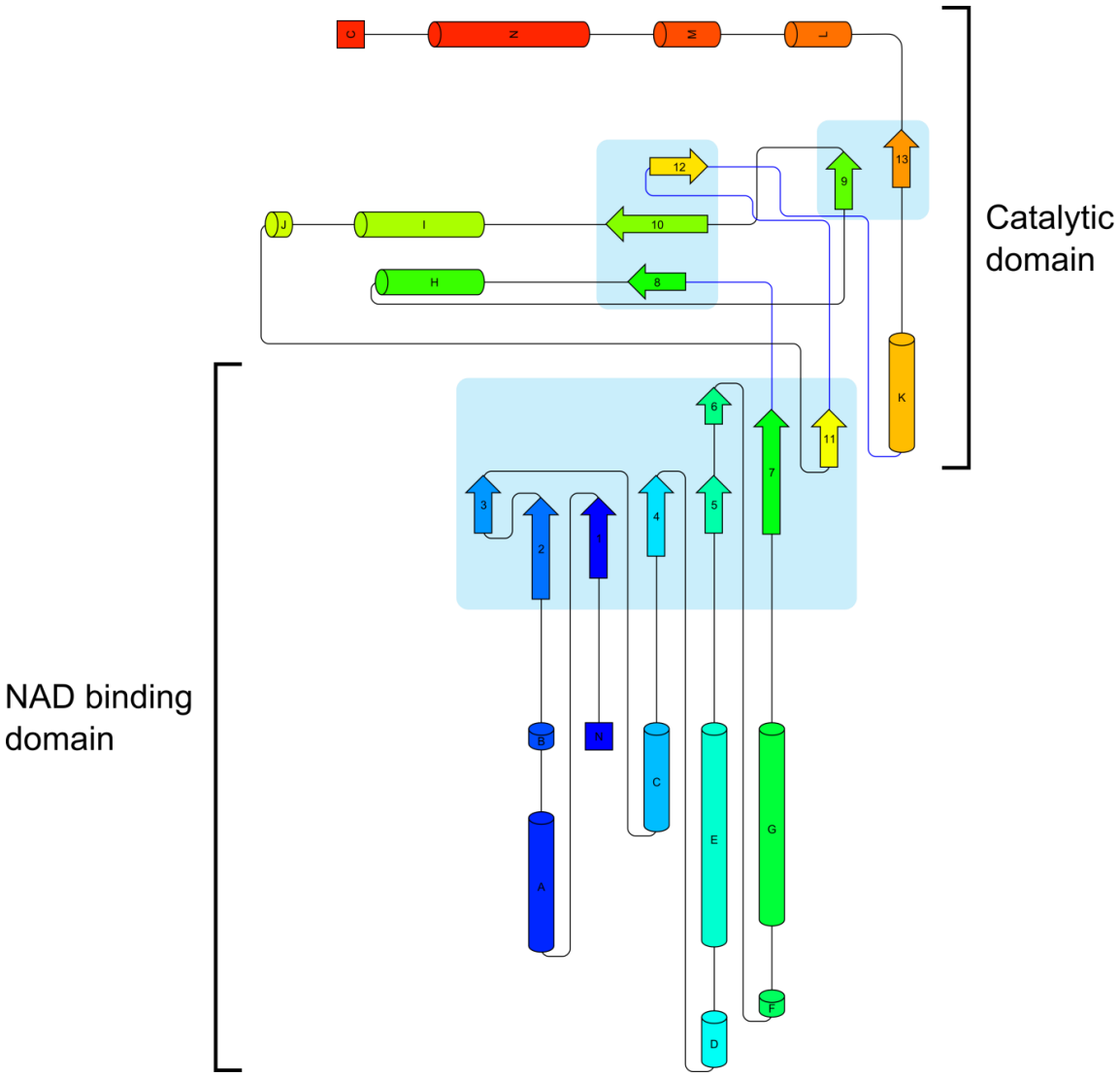


Figure 3.1.12 - protein topology diagram of TDH. This diagram was drawn using Pro-origami(263) and a monomeric TDH structure model from TK8. The diagram was drawn using the Dictionary of Secondary Protein Structure (DSSP) to assign secondary structure elements(264,265). α-helices are represented by barrels and assigned a letter in alphabetical order. B-strands are represented by flat arrows and are numbered in ascending order. The secondary structure elements are coloured from blue, starting at the N-terminus, through green, yellow and orange to red at the C-terminus.

TDH possesses a pattern of parallel β-sheets, which are flanked by α-helices. This motif, termed the Rossmann fold is common to NAD-binding proteins(266,267), and hence SDRs(118,261). The protein topology diagram in Figure 3.1.12 shows how the sequence of alternating β-strands and α-helices arrange so that the β-strands are parallel and spatially adjacent. The 3D orientation of NAD in relation to the Rossmann fold can be seen in Figure 3.1.13.

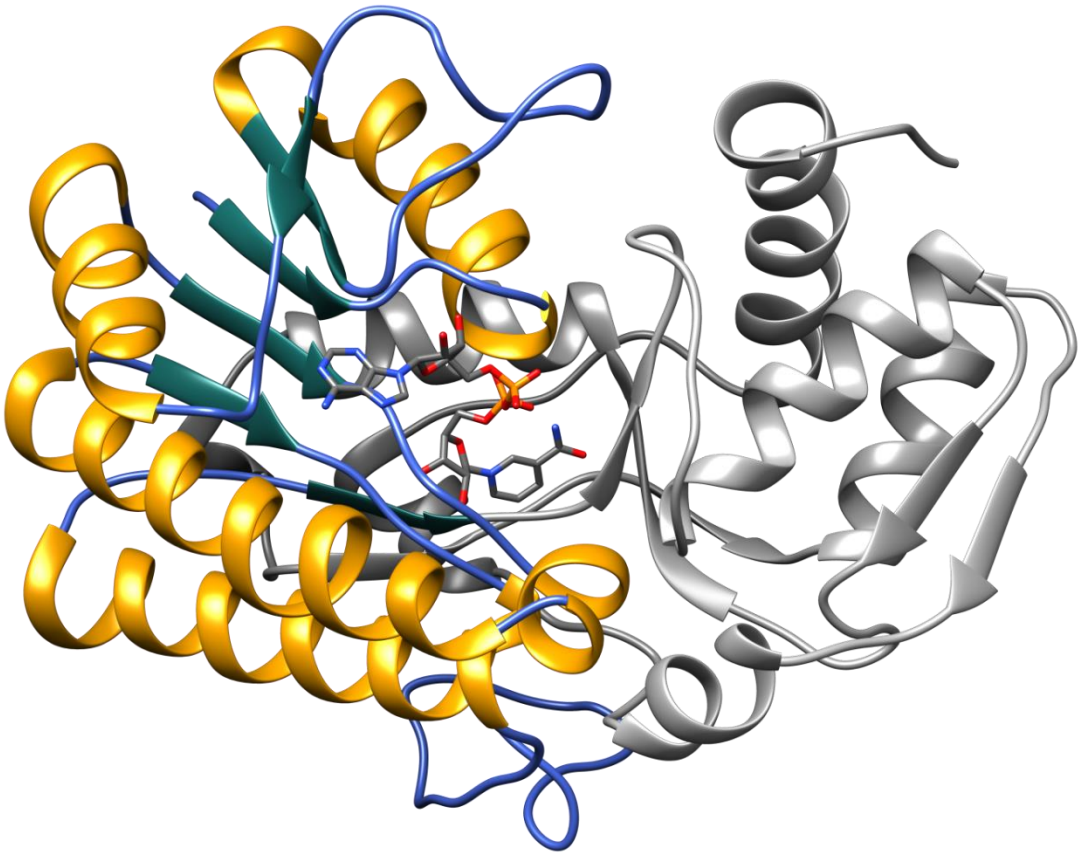


Figure 3.1.13 - ribbon representation of a TDH monomer. The NAD-binding Rossmann fold is coloured according to secondary structure, whilst the remainder of the molecule is coloured grey. The Rossmann fold is characterised by a six-stranded parallel β-sheet that is flanked by α-helices.

The structural motif GxxGxxG, which is common to the 'extended' class of SDRs (eSDR), is located within the NAD binding region and is represented by the sequence Gly9-Ala10-Leu11-Gly12-Gln13-Ile14-Gly15. The motif appears to be involved in binding the diphosphate group of NAD. Simultaneously, the YxxxK motif and several other conserved residues are located in the active sites of SDRs. TDH from *T. brucei* is no exception. These residues are displayed in Figure 3.1.14. They include Met81, Ser82, Thr119, Thr186, Trp280, and Tyr144 and Lys148 of the eSDR YxxxK motif.

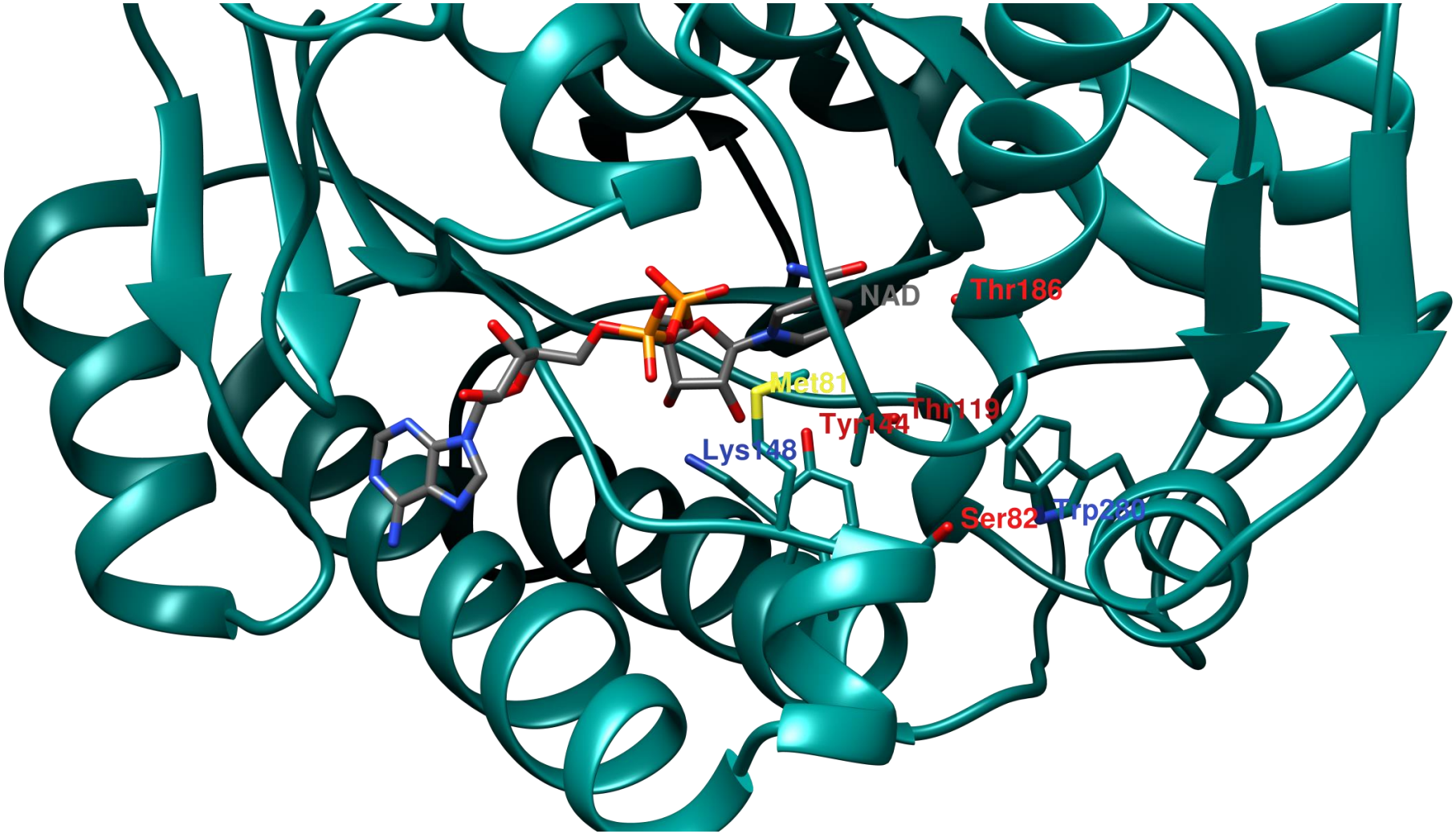


Figure 3.1.14 - ribbon representation of TDH, showing the conserved active site residues, Met81, Ser82, Thr119, Tyr144, Lys148, Thr186 and Trp280.

Although no large domain movements were observed between the various TDH structure models, aligning different monomers from various models showed that there was a certain degree of variation in the relative positions of different atoms (see Section 2.1.3.3 for methods). The fact that different areas of the protein showed more variation than others suggested that there are regions within the protein that are more flexible. This was illustrated clearly by observing the RMSD of α-carbon positions in the aligned residues. Another indication of flexibility or disorder is given by looking at the average B-factor for each residue. When the B-factors are averaged across different structure models, one can see which areas of the protein tend to be more disordered. Figure 3.1.15 to Figure 3.1.18 inclusive provide a visual description of the flexibility in TDH structure.

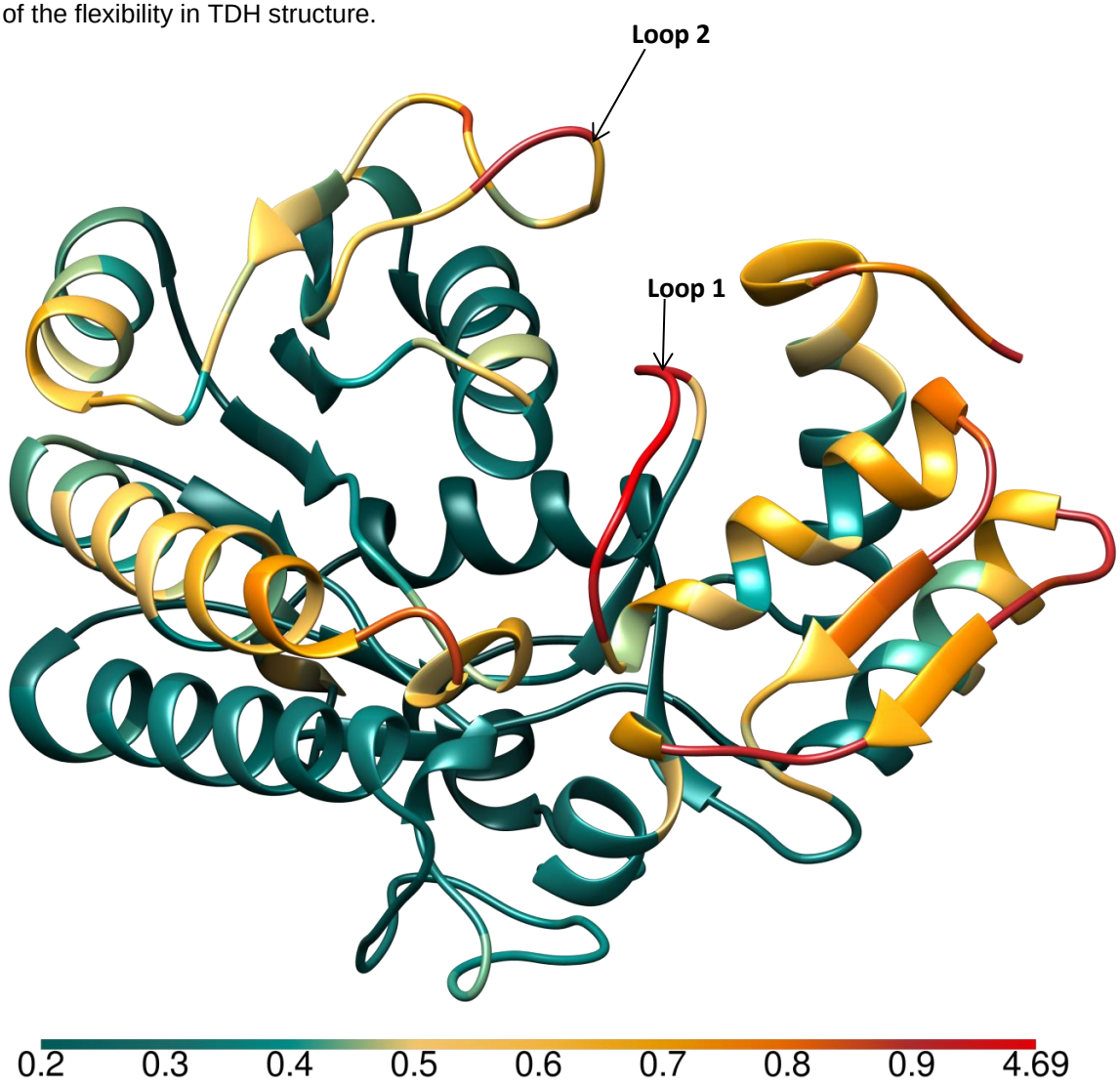


Figure 3.1.15 - ribbon representation of TDH, coloured by the RMSD of α-carbon positions in each residue. The RMSD was determined after aligning eight representative monomeric TDH models in UCSF Chimera(182). The scale at the bottom of the diagram indicates the colour scheme relating to the RMSD measured in Angstrom.

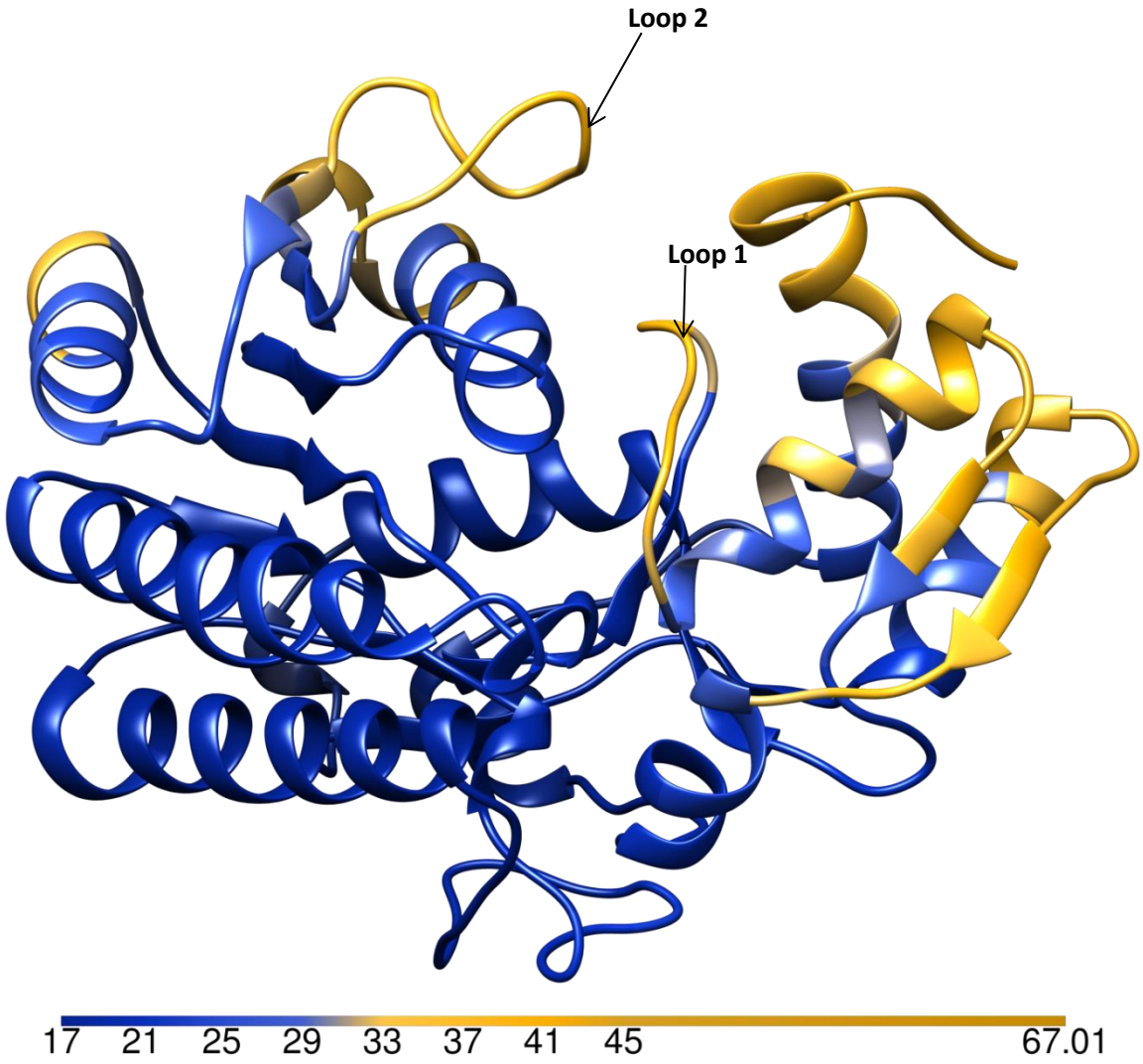


Figure 3.1.16 - ribbon representation of TDH, coloured by the average B-factor of each residue. The average B-factor per residue was calculated for eight representative monomeric TDH models using the CCP4 program baverage(268). The mean of all the average B factors was then determined from these values. The scale at the bottom of the diagram indicates the colour scheme relating to the B-factor, measured in Angstrom squared.

Looking at Figure 3.1.15 and Figure 3.1.16, there are number of regions that appear to be flexible. Firstly, there are two disordered loop regions: Loop 1 is between Thr179 and Ala185, and Loop 2 lies between residues 35 and 60. There also appears to be a high degree of flexibility in the catalytic domain (on the right-hand side of Figure 3.1.15 and Figure 3.1.16). The bar charts below highlight the positions along the TDH sequence that correspond to regions of flexibility.

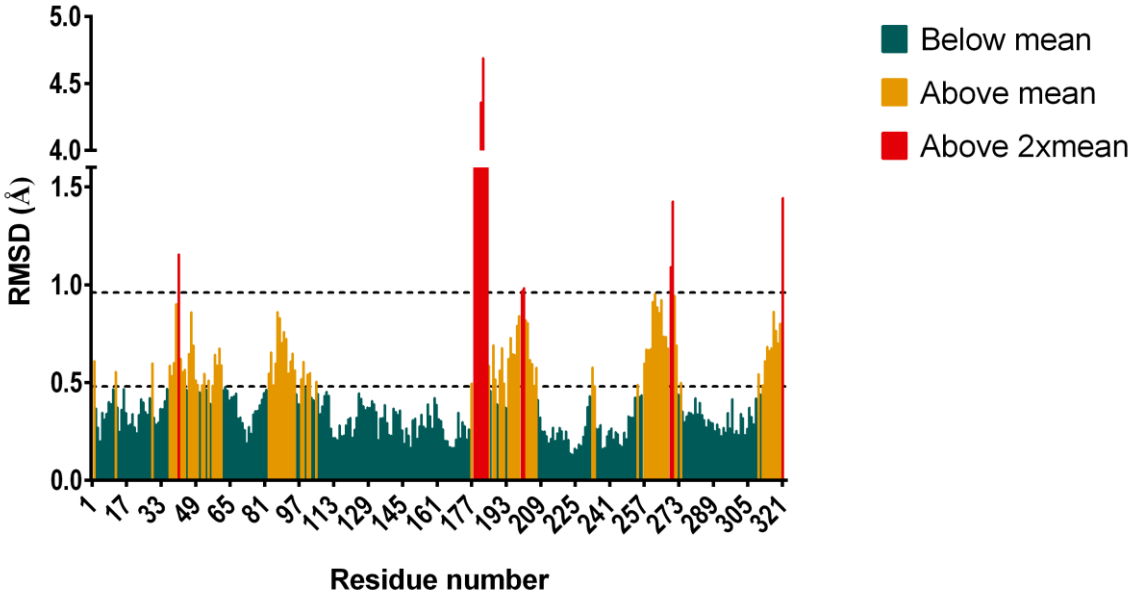


Figure 3.1.17 - bar chart showing the RMSD of α-carbons per TDH residue. The lower horizontal dotted line represents the mean RMSD value of all residues. Bars representing values higher than this are coloured gold. The upper dotted horizontal line is double the mean RMSD, and bars above this line are coloured red. The RMSD was determined after aligning eight representative monomeric TDH models in UCSF Chimera(182).

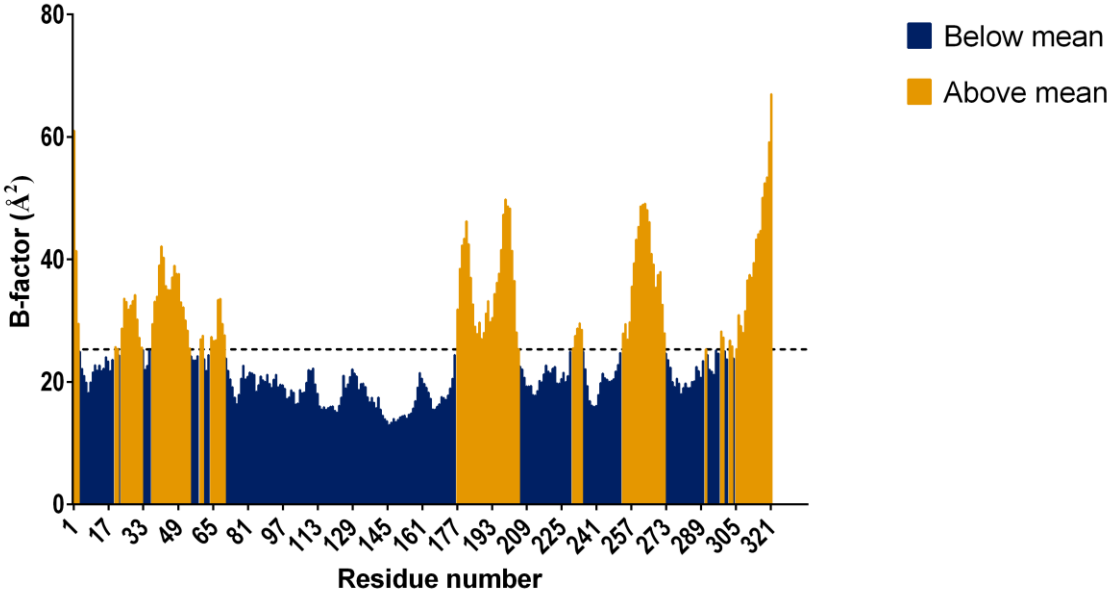


Figure 3.1.18 - bar chart showing the average B-factor per TDH residue. The dotted horizontal line indicates the mean B-factor across all residues. Bars representing values higher than this line are coloured gold. The average B-factor per residue was calculated for eight representative monomeric TDH models using the CCP4 program baverage(268). The mean of all the average B factors was then determined from these values.

The same flexible or variable regions are highlighted by comparison of average B-factors, as by comparison of RMSDs. There appears to be flexibility in the NAD-binding region, particularly around Loop 2. Although the B-factors are low, the RMSDs around residues 82 to 100 are particularly high. This region lies at the dimerization interface (discussed below), indicating some conformational variability there. The most striking evidence of conformational variability is found at the residues of Loop 1, Thr179 to Ala185. The significance of Loop 1 and its different positions will be discussed in more detail below. The region adjacent to Loop 1, including residues up to Pro205, is highly disordered and shows a lot of variation between TDH models. In three dimensions, this region can be seen as one of a pair of adjacent α-helices and β-strands, which can be seen on the right hand sides of the structures in Figure 3.1.15 and Figure 3.1.16. The second disordered α-helix-β-sheet in this region occurs around residues 255-273, which also stands out as a disordered region in Figure 3.1.15 and Figure 3.1.16.

Loop 1 was the region that showed the largest apparent change in its conformation (see Figure 3.1.17). The loop actually showed two distinctive positions, one where the loop was in an 'open' position, lying at a distance from the active site, and another where the loop is 'closed' and lies adjacent to the active site. Therefore, for each model of a TDH monomer, the structure can be classified as 'open' or 'closed', on the basis of the conformation of Loop 1. The structure models used in the analyses that produced Figure 3.1.15 to Figure 3.1.18 included structures in which Loop 1 lay in an open conformation and others in which Loop 1 was closed. Comparing the average B-factors in open and closed structures can highlight any changes in flexibility or

disorder that accompany the change in the Loop 1 conformation. Figure 3.1.19 shows the average B-factors of TDH residues in 'open' structures, whilst Figure 3.1.20 shows the average B factors of TDH residues in 'closed' structures.

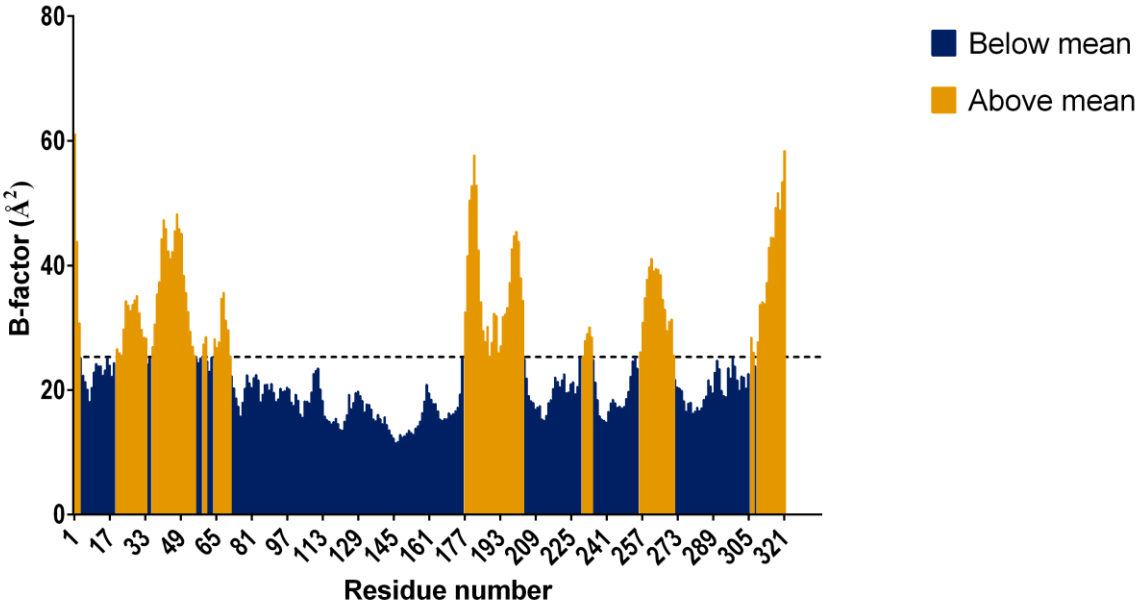


Figure 3.1.19 - bar chart showing the average B-factor per TDH residue, taken as a mean from the average B factors of residues in four of the eight representative monomers used in these analyses. In each structure, Loop 1 is in the 'open' conformation.

In the 'open' structures, the highest B-factors are seen around the loop regions. However, in the closed structures, although these same residues are above the mean all-residue B-factor, they are much less disordered. The catalytic domain appears to be much more disordered, and residues 186-207 and residues 251-274 have particularly high B-factors. The residues from Gln296 leading up to the C-terminus are also very disordered. This was also observed during real space model refinement; these disordered regions were associated with less difference density.

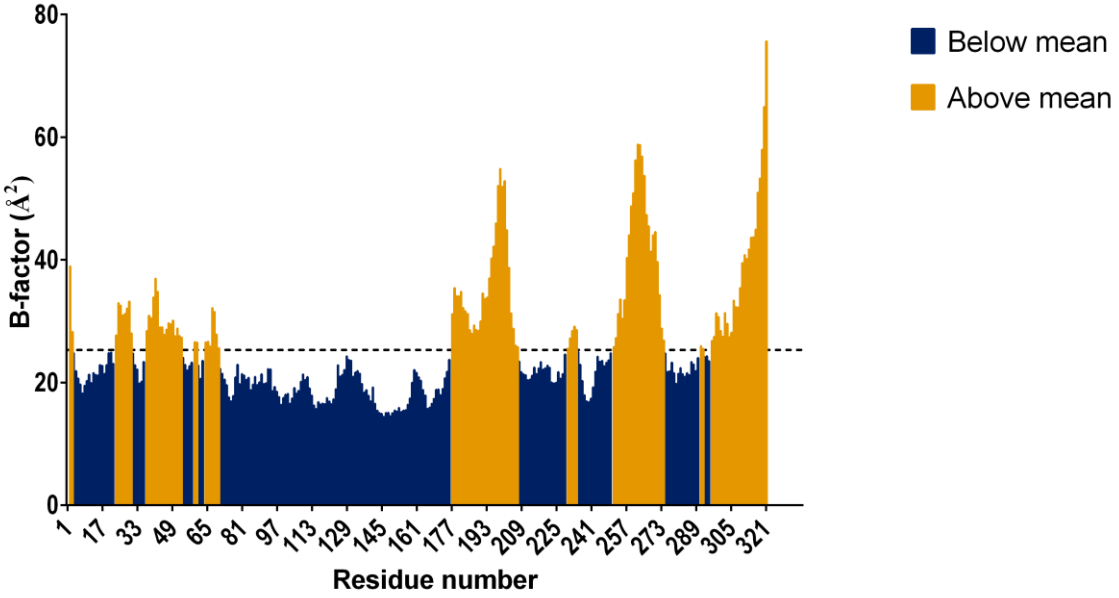


Figure 3.1.20 - bar chart showing the average B factor per TDH residue, taken as a mean from the average B factors of residues in four of the eight representative monomers used in these analyses. In each structure, Loop 1 is in the 'closed' conformation.

These findings indicate that there is a relationship between the conformation of Loop 1, and the flexibility of other regions of the protein, particularly the catalytic domain. Three of the four 'closed' structures examined were bound to NAD/NADH and another ligand, whilst only one of the four 'open' structures examined was bound to NAD and another ligand. The different conformational changes do not appear to correlate with the space-group of the crystal from which the data was collected. Thus, the specific conformation of Loop 1 and the flexibility of the catalytic region may actually be influenced by ligand binding.

As discussed above, the change from an open conformation of Loop 1 to a closed conformation has a dramatic effect on the nature of the active site, and so warrants closer examination.

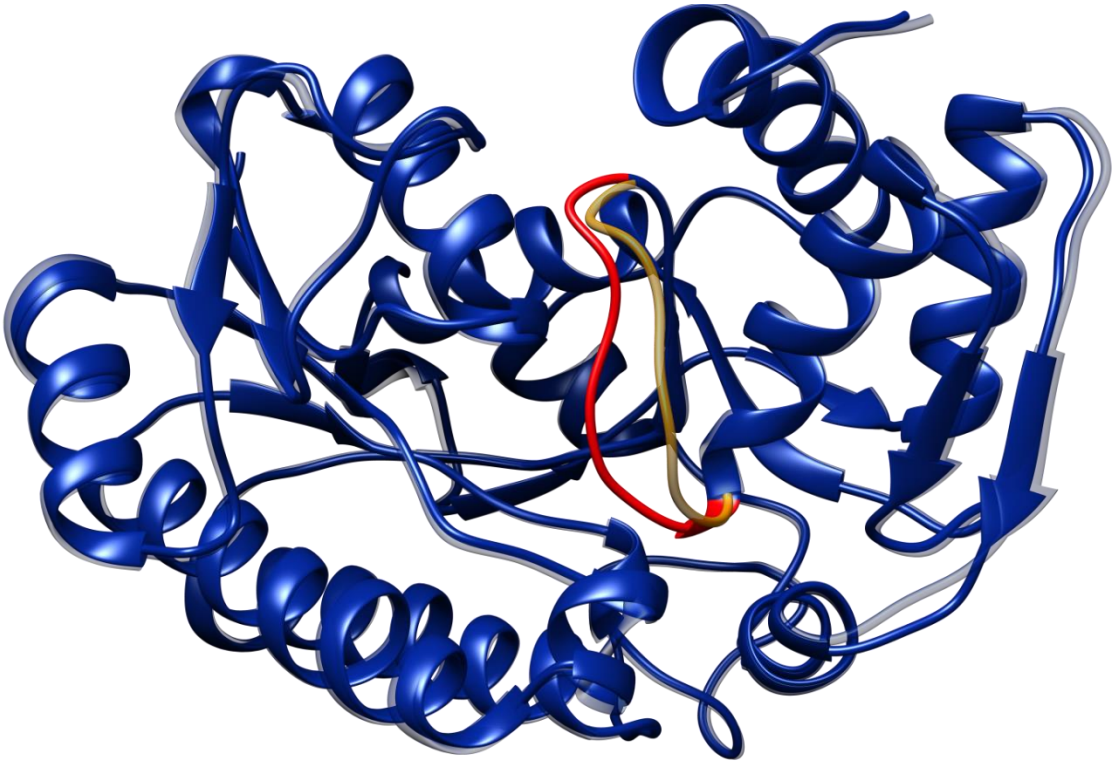


Figure 3.1.21 - ribbon representations of two superposed monomeric TDH structures. One structure is the open conformer and is semi-transparent. The other structure is the closed conformer, and this is opaque. For emphasis, Loop 1 of the open conformer is coloured gold, whilst Loop 1 of the closed conformer is coloured red. The main difference between the two conformations manifests as a closure by Loop 1 over the active site.

The NAD binding site and the L-threonine binding site lie adjacent to each other and are connected. The side chain of Met81 appears to act as a barrier between the two active sites and may help in positioning the two substrates on their initial approach to the active site. Although the L-threonine binding site is much smaller than the NAD binding site, it is open to the exterior of the protein, so it could theoretically accommodate molecules that are larger than L-threonine. However, when Loop 1 is in its closed conformation, the active site is completely sealed off, leaving little room for a molecule much bigger than L-threonine. This has clear implications for drug discovery, as will be discussed further.

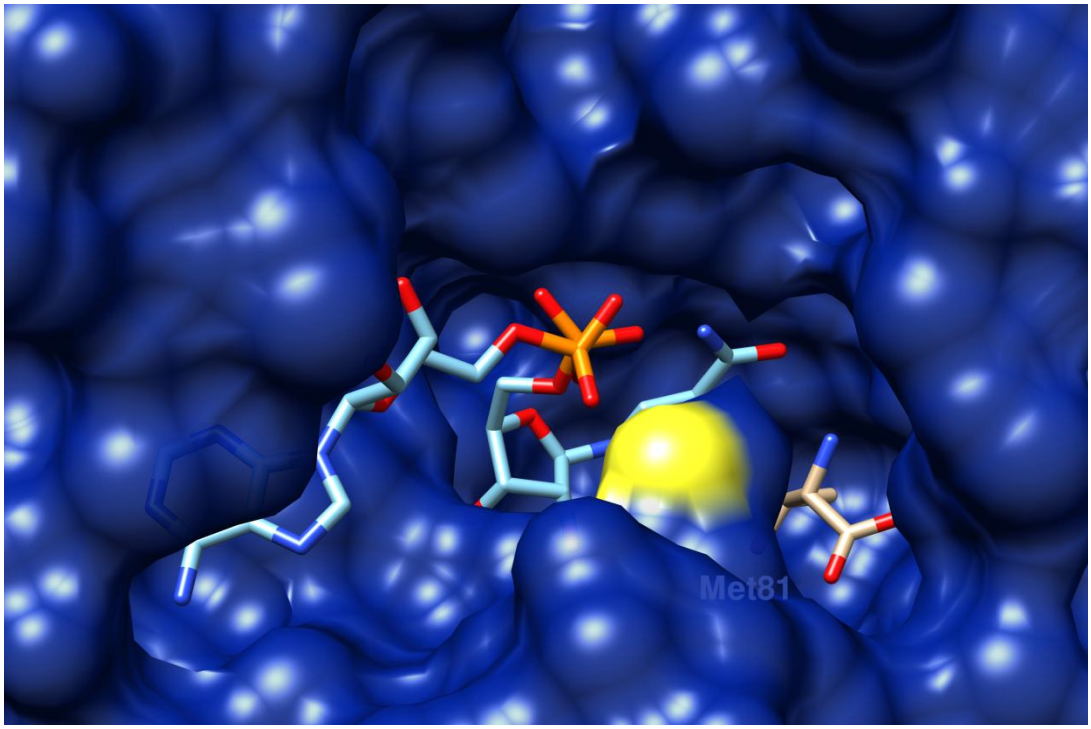


Figure 3.1.22 - surface representation of TDH with NAD^+ and L-threonine bound. The image was created by superposing an NAD^+-bound monomer from model TK8 and a NADH- and L-threonine-bound TDH monomer from model TNADH4. The open conformer is shown here. It can be seen that Met81 forms a partial barrier between the binding pockets of each substrate.

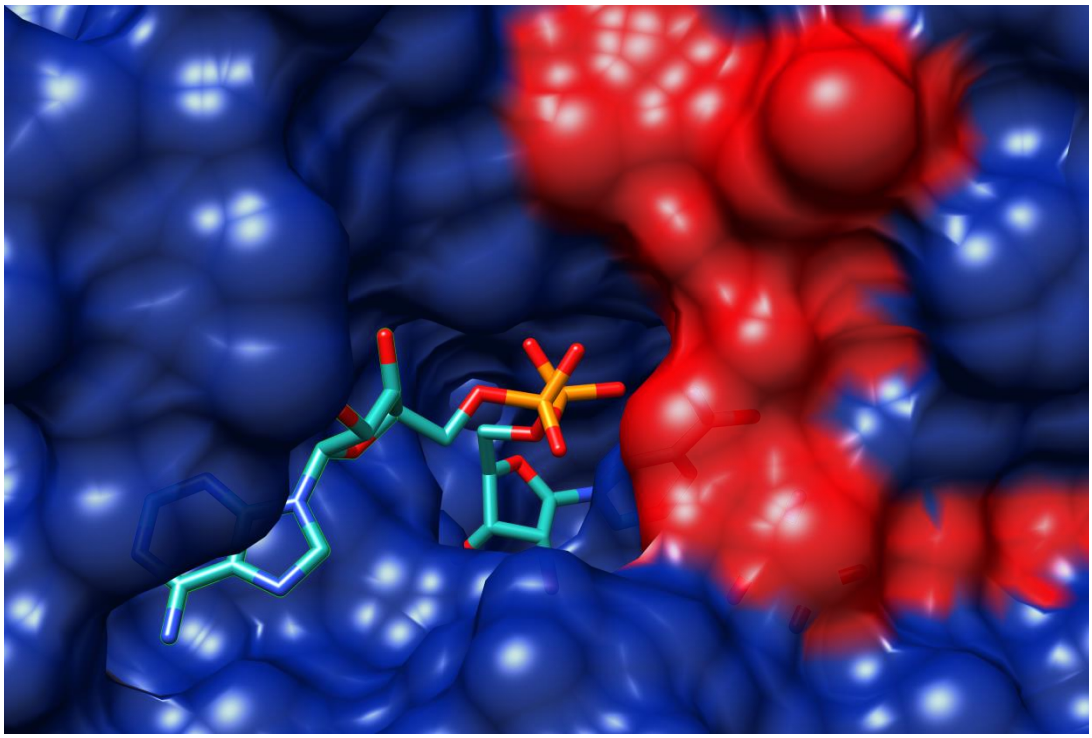


Figure 3.1.23 - surface representation of TDH with NAD^+ and L-threonine bound. The image was created by superposing an NAD^+-bound monomer from model TK8 and a NADH- and L-threonine-bound TDH monomer from model TNADH4. The closed conformer is shown here. Loop 1 is coloured red to highlight the fact that it seals off the TDH binding site from the external environment.

3.1.6 Quaternary structure & Complex Formation

The studies described in this thesis provided several lines of evidence in support of TDH existing as a homodimer. Below, evidence from X-ray crystallography, size-exclusion chromatography and cross-linking studies are described.

TDH was revealed to exist as a dimer during solution and refinement of X-ray crystallographic models. When carrying out real space refinement using the software Coot, by default only the molecules in the asymmetric unit are shown. Therefore, symmetry related molecules are hidden, so only a monomer would be visible sometimes. In such cases the existence of the dimer was easily confirmed by selecting the option to display symmetry-related molecules. Further evidence of the existence of the dimer, and evidence that this was the largest oligomer present, was provided by Protein Interfaces, Surfaces and Assemblies (PISA) analyses.

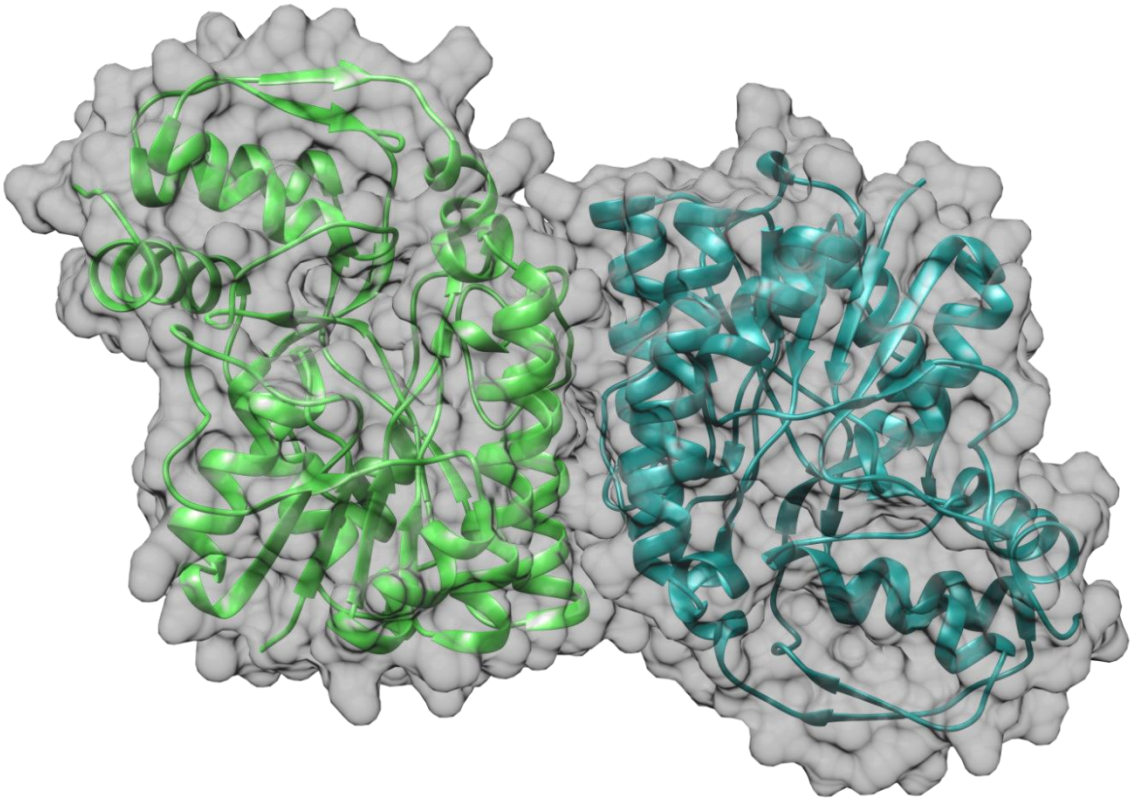


Figure 3.1.24 - ribbon and surface representation of the TDH homodimer.

The dimerization interface consists of two α-helices (Pro89-Tyr110 and Val143-Tyr162) and a loop (lys126-Thr142), and the same regions on each monomer make contact in the dimer. As can be seen in Figure 3.1.25, the dimer is stabilised by a large number of hydrophobic, polar and ionic interactions. In particular, polar side chains of residues on each monomer form hydrogen bonds with each other and charged residues form salt bridges. This probably gives greater stability to the dimer, whilst the length of some of the side chains allows their interactions to be maintained in the case of any change in the orientation of the two monomers.

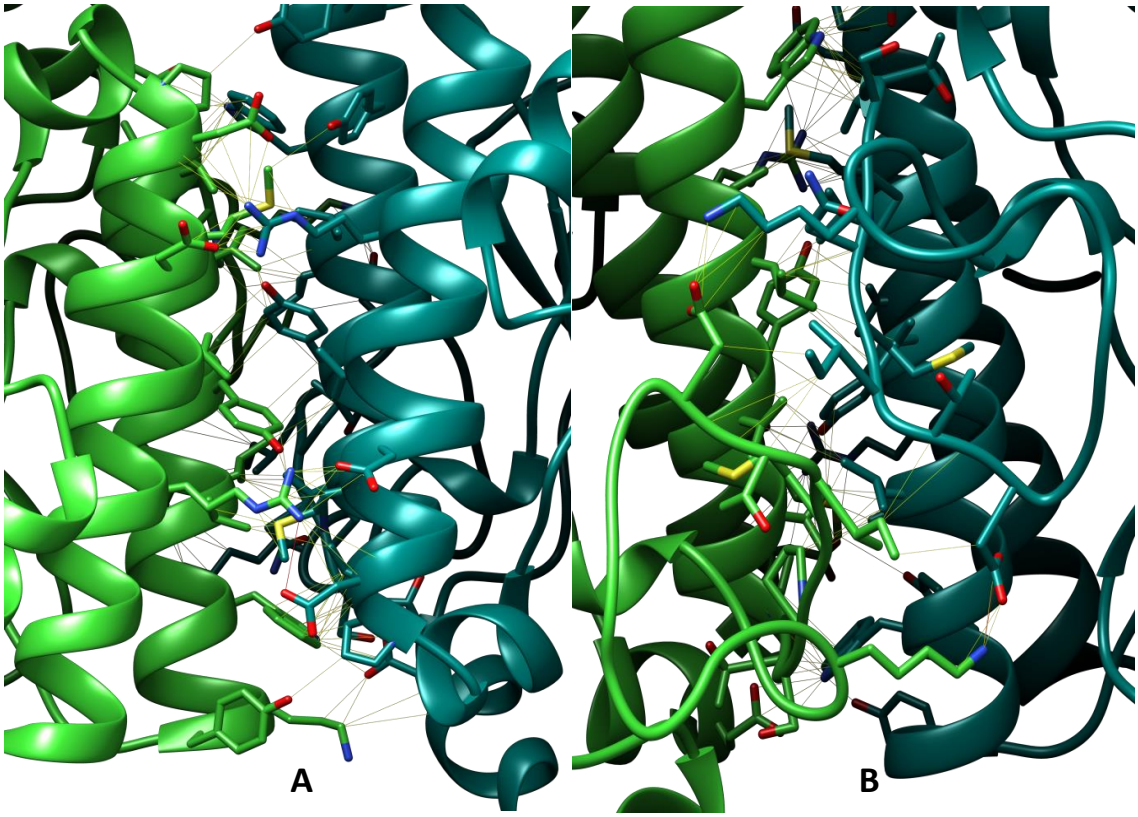


Figure 3.1.25 - the TDH dimerisation interface. The protein is represented by a ribbon and important interacting amino acids are shown in the stick representation. Yellow lines represent potential contacts (when Van der Waals radii of nearby atoms overlap). Red lines indicate atoms within hydrogen bonding distance. The images labelled A and B show the interface from opposite angles.

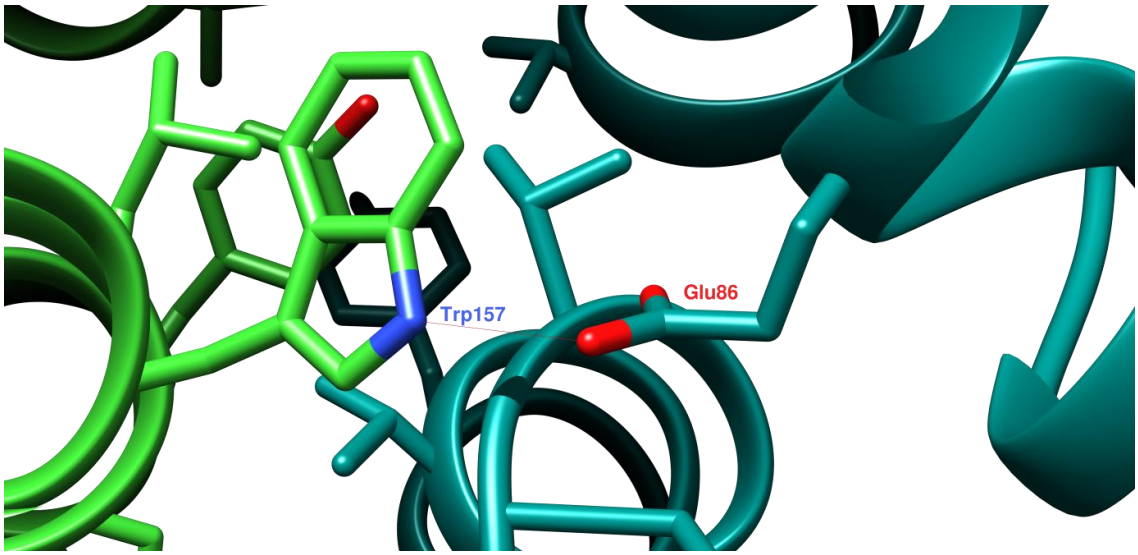


Figure 3.1.26 - the interaction between Glu86 and Trp157 appears to be one of the hydrogen bonds stabilising the dimer.

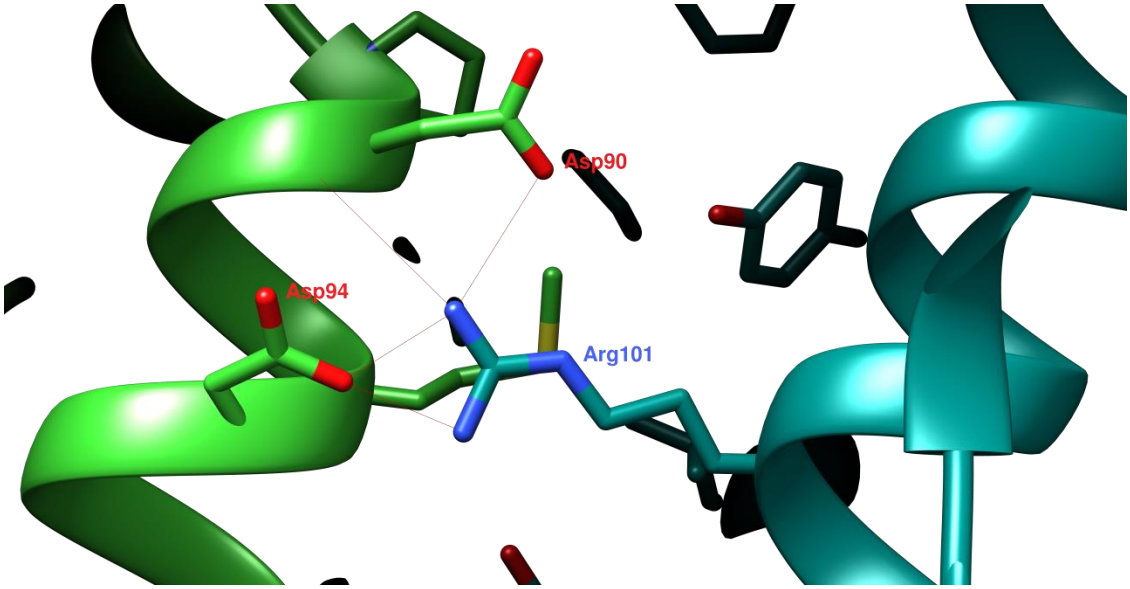


Figure 3.1.27 - the interaction between Asp90 and Asp94 and Arg101 appears to be a group of salt bridges stabilising the dimer. The length and flexibility of Arg101, could help to maintain the interaction upon movement of the subunits of the dimer.

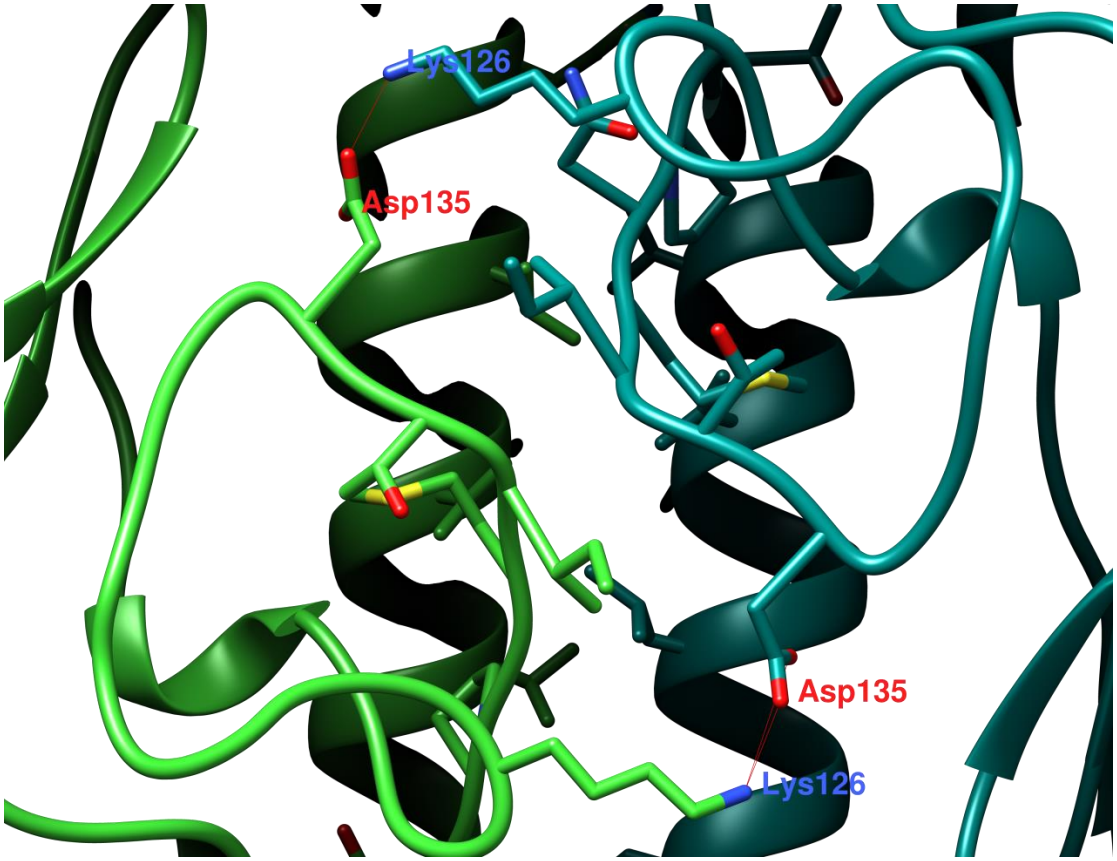


Figure 3.1.28 - Lys126 and Asp135 appear to be involved in another salt bridge stabilising the dimer.

As is evident in Table 3.1.1, TDH models were derived from crystals with different characteristics. Therefore, an analysis of the relationships between different subunits was carried out using the CCP4 software, PISA, to supplement the findings described above. Table 3.1.3 lists a number of different interfaces between different TDH subunits in the different structure models. The only stable association between TDH subunits was the dimer relationship already described. PISA detected one possible multimer involving four subunits in the model TNADH4, but could not determine whether this would be stable in solution. Therefore, another CCP4 program, pdbset, was used to generate a .pdb file containing all symmetry-related molecules in addition to those in the asymmetric unit. As can be seen in Figure 3.1.29, the only quaternary structure evident is the TDH dimer. The buried surface area was similar for the dimerisation interfaces from different crystals. Differences in the buried surface area result from conformational changes around the dimer interface, as shown in Figure 3.1.15 and Figure 3.1.17 (see residues 89-100). There was no clear relationship between the nature of the dimerisation interface and the space group, TDH conformation (open or closed) and whether TDH was ligand bound.

Table 3.1.3 - TDH quaternary structure identified by PISA and confirmed visually. * ΔG_{diss} = dissociation barrier.

Model	Oligomer	Accessible surface area ($Å^2$)	Buried surface area ($Å^2$)	ΔG_{diss}^* ($kcal.mol^{-1}$)	Stable in solution?	Visual confirmation
Model: TK8 **Space group:** $P2_12_12$ **Unit cell parameters:** a=132.04, b=276.49, c=55.74; αβγ=90° **Molecules in the asymmetric unit:** 6	Dimer	24918.1	5470.9	3.3	Yes	✓
	Dimer	24832.7	5449.2	2.6	Yes	✓
	Dimer	24993.7	5428.6	2.3	Yes	✓
Model: TS2 **Space group:** $P4_32_12$ **Unit cell parameters:** ab=91.76, c=93.60; αβγ=90° **Molecules in the asymmetric unit:** 1	Dimer	25239.1	5028.6	3.2	Yes	✓

Model	Oligomer	Accessible surface area ($Å^2$)	Buried surface area ($Å^2$)	ΔG_{diss}^{*} (kcal.mol^{-1})	Stable in solution?	Visual confirmation
Model: TIO8 **Space group:** $P2_12_12$ **Unit cell parameters:** a=90.40, b=131.53, c=55.02; αβγ=90° **Molecules in the asymmetric unit:** 2	Dimer	24519.8	4880.8	4.4	Yes	✓
Model: TQ5 **Space group:** $P2_12_12$ **Unit cell parameters:** a=133.03,b=273.06, c=55.80; αβγ=90° **Molecules in the asymmetric unit:** 2	Dimer	24264.3	7309.4	4.1	Yes	✓
Model: AAT1 **Space group:** $P2_12_12$ **Unit cell parameters:** a=133.03, b=273.06, c=55.80; αβγ=90° **Molecules in the asymmetric unit:** 6	Dimer	24167.7	4813.6	4.3	Yes	✓
	Dimer	24294.1	4772.0	3.1	Yes	✓
	Dimer	24217.8	4790.1	2.4	Yes	✓

Model	Oligomer	Accessible surface area ($Å^2$)	Buried surface area ($Å^2$)	ΔG_{diss}^{*} (kcal.mol^{-1})	Stable in solution?	Visual confirmation
Model: TNADH4 **Space group:** $P2_12_12$ **Unit cell parameters:** a=83.45, b=136.12, c=55.69; αβγ=90° **Molecules in the asymmetric unit:** 2	Dimer	24587.0	5788.1	0.3	Yes	✓
	Tetramer	46831.2	13919.2	-1.6	Uncertain	x (see Figure 3.1.29)
Model: TM131 **Space group:** $P2_12_12$ **Unit cell parameters:** a=133.45, b=278.63, b=56.27; αβγ=90° **Molecules in the asymmetric unit:** 6	Dimer	24779.8	6050.2	4.5	Yes	✓
		24400.6	6351.1	4.1	Yes	✓
		24521.6	6357.1	3.1	Yes	✓

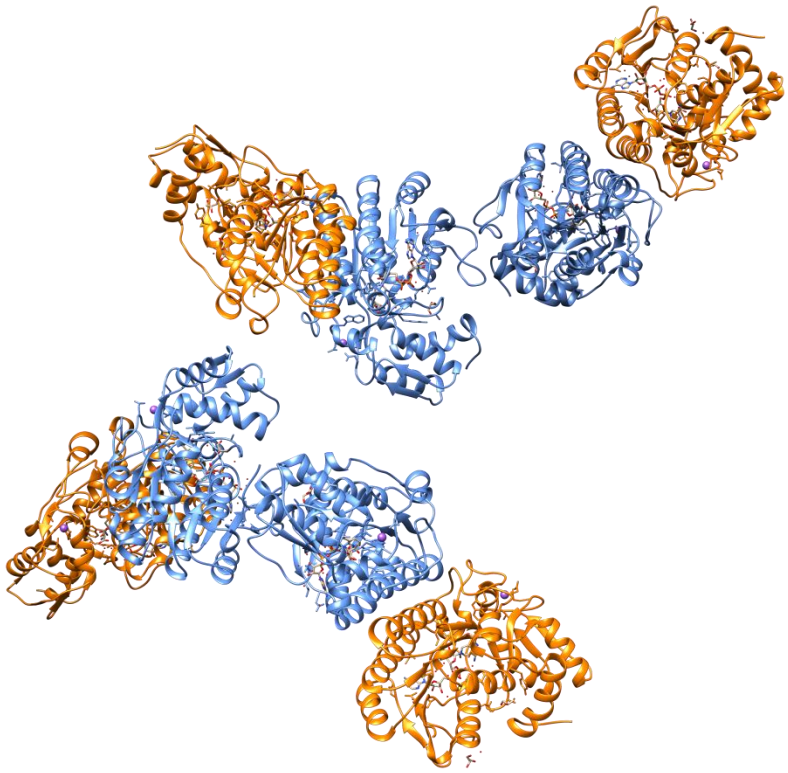


Figure 3.1.29 - visualisation of TDH structure TNADH4, from the .pdb file generated by pdbset. All symmetry-related molecules are present, in addition to those in the asymmetric unit.

Further information on the quaternary structure of TDH and KBL was gained through size-exclusion chromatography experiments, as described in Section 2.1.4.1. The elution volumes of TDH during size-exclusion chromatography also provided evidence that TDH is a homodimer. The elution volumes were found to be different under different conditions, suggesting that the quaternary structures of TDH and KBL, which were also studied by this method, were influenced by the presence of other proteins and substrates. The results of a calibration of the gel filtration column with protein standards are shown in Figure 3.1.30. In Table 3.1.4, the elution volumes of TDH, KBL and mixtures of both are listed, along with the molecular weights that those elution volumes correspond to, as calculated by interpolation from the standard curve (see Figure 2.1.10). The fractions corresponding to peaks recorded during elution from the column were all analysed using SDS PAGE to confirm the presence of protein in that fraction. NAD^+ caused very large peaks when eluting from the column (elution volume ~24ml), so the presence of protein in the corresponding fractions was excluded using SDS-PAGE and silver staining in those cases.

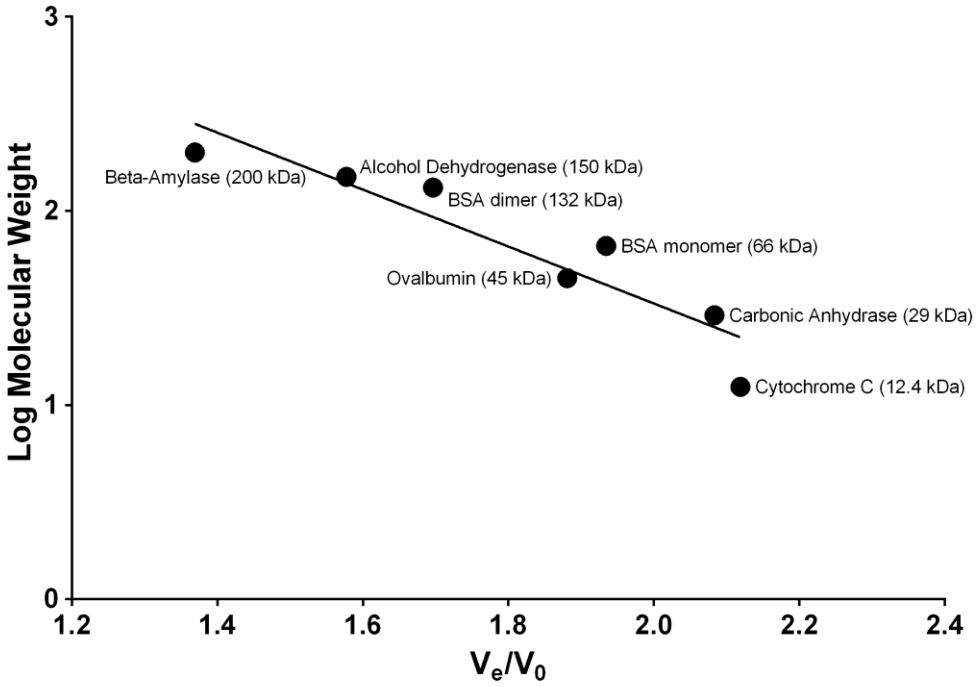


Figure 3.1.30 - calibration of the size-exclusion chromatography column using standard proteins. The log molecular weights of the proteins are plotted as a function of their elution volume (V_e) divided by the void volume (V_0). The void volume was determined by measuring the elution of blue Dextran, which eluted at 8.4ml.

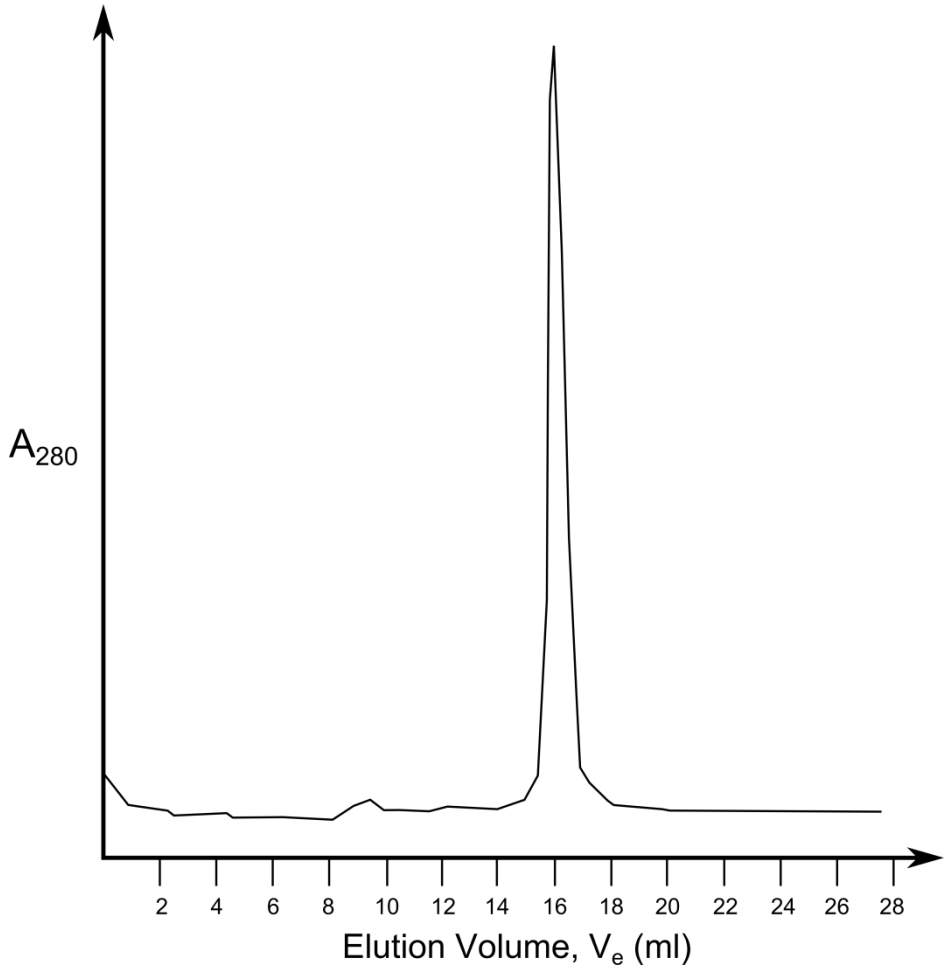


Figure 3.1.31 - an example of a chromatogram from one of the size-exclusion chromatography experiments. This chromatogram is of a sample containing TDH at 1.1 x 10^{-1} mM and 30 mM L-threonine.

Figure 3.1.31 shows an example of a chromatogram, from which the elution volumes corresponding to absorbance peaks were taken. Table 3.1.4 shows the various samples tested in size-exclusion chromatography experiments, together with the elution volumes and estimated molecular weights.

Table 3.1.4 - Results of size-exclusion chromatography experiments. Note that some results appear more than once, under different headings.

Sample Contents	Ve (ml)	Ve/V_0	Interpolated MW (kDa)	Multiple of MW	Suggested Oligomeric State
TDH					
TDH (1.1 $x10^{-1}$ mM [4mg/ml])	15.5	1.85	56.2	1.49	Monomer/Dimer
TDH (1.8 $x10^{-1}$ mM [6.8mg/ml])	15	1.79	68.8	1.82	Dimer
TDH (1.1 $x10^{-1}$ mM [4mg/ml]) + 10mM NAD^+	15	1.79	68.8	1.82	Dimer
TDH (1.1 $x10^{-1}$ mM [4mg/ml]) + 30mM L-threonine	16	1.90	46.0	1.22	Monomer
TDH (1.1 $x10^{-1}$ mM [4mg/ml]) + 10mM NAD^+ + 20mM L-threonine	16.5	1.96	37.6	0.99	Monomer
TDH (1.1 $x10^{-1}$ mM [4mg/ml]) + 10mM NAD^+ + 30mM L-threonine	16.1	1.92	46.0	1.22	Monomer
TDH (4.8 $x10^{-2}$ mM [1.8mg/ml]) + KBL (4.4 $x10^{-2}$ mM [2mg/ml]) + 10mM NAD^++ 30mM L-threonine	16.5	1.96	37.6	0.99	Monomer
TDH (5.3 $x10^{-2}$ mM [2mg/ml]) + 30mM L-allo-threonine	14	1.67	102.8	2.72	Trimer
TDH (5.3 $x10^{-2}$ mM [2mg/ml]) + 10mM NAD^+ + 30mM L-allo-threonine	15	1.79	68.8	1.82	Dimer
TDH (5.3 $x10^{-2}$ mM [2mg/ml]) + 60mM pyruvate	15.2	1.81	63.4	1.68	Dimer
KBL					
KBL (4.4 $x10^{-2}$ mM [2mg/ml])	14.4	1.71	87.5	1.91	Dimer
KBL (6.54 $x10^{-2}$ mM [3mg/ml]) + 20mM Gly	11	1.31	343.5	7.48	Octamer

Sample Contents	Ve (ml)	Ve/V_0	Interpolated MW (kDa)	Multiple of MW	Suggested Oligomeric State
KBL (4.4 $x10^{-2}$ mM [2mg/ml]) + 1mM PLP	14.5	1.73	84.0	1.83	Dimer
TDH (4.8 $x10^{-2}$ mM [1.8mg/ml]) + KBL (4.1 $x\ 10^{-2}$ mM [1.9mg/ml]) + 10mM NAD^+ + 20mM L-threonine + 1mM PLP + 1mM CoA	16.5	1.96	37.6	0.82	Monomer
TDH + KBL					
TDH (5.3 $x10^{-2}$ mM [2mg/ml]) + KBL (4.4 $x10^{-2}$ mM [2mg/ml]) - first peak	12.5	1.49	187.7	4.97 (TDH); 4.09 (KBL)	Multi-enzyme complex
TDH (5.3 $x10^{-2}$ mM [2mg/ml]) + KBL (4.4 $x10^{-2}$ mM [2mg/ml]) - second peak	16.8	1.99	34.0	0.90 (TDH); 0.74 (KBL)	Monomers (TDH and/or KBL)
TDH (4.8 $x10^{-2}$ mM [1.8mg/ml]) + KBL (4.4 $x10^{-2}$ mM [2mg/ml]) + 10mM NAD^++ 30mM L-threonine	16.5	1.96	37.6	0.99 (TDH); 0.81 (KBL)	Monomers (TDH and/or KBL)
TDH (4.8 $x10^{-2}$ mM [1.8mg/ml]) + KBL (4.1 $x\ 10^{-2}$ mM [1.9mg/ml]) + 10mM NAD^+ + 20mM L-threonine + 1mM PLP + 1mM CoA	16.5	1.96	37.6	0.99 (TDH); 0.81 (KBL)	Monomers (TDH and/or KBL)

In several experiments involving TDH, the elution volume of the protein corresponded to a protein of molecular weight 63-68 kDa. This suggests that TDH was present in its dimeric form, which would have a molecular weight around 74 kDa. A very surprising finding was that TDH samples containing L-threonine eluted at a volume that corresponded to the molecular weight of a TDH monomer. This raises the possibility that L-threonine binding or catalytic activity promotes the dissociation of the TDH dimer into monomers. Samples containing NAD^+ and the TDH inhibitors L-*allo*-threonine and pyruvate eluted at volumes corresponding to the TDH dimer, adding to the likelihood that catalysis causes the increase in elution volume. There is one outlier, which had a higher elution volume corresponding to a protein of approximately 1.5 times the molecular weight of a TDH monomer (56.2 kDa). When compared with other findings, it appears likely that this corresponds to the TDH dimer. A second outlier resulted from an experiment with a mixture of TDH and L-*allo*-threonine, without added NAD^+. The elution volume of TDH in this sample suggests a molecular weight corresponding to a trimer of TDH. Further investigations will be required to ascertain whether this is an erroneous result, or if there is another higher oligomer of TDH. A sample containing a mixture of TDH and KBL together with NAD^+ and L-threonine eluted together as one broad peak. The elution volume at the peak of the tip suggests that the proteins had a molecular weight of 37.6 kDa, suggesting that monomers of TDH and KBL co-eluted and that the two proteins did not form a complex.

Judging by the elution volumes of samples of KBL, it is highly probable that this protein exists as a dimer. An elution volume of 84 kDa was measured which corresponds to a KBL dimer. The addition of pyridoxal phosphate (PLP) seemed to have no effect on the elution volume of KBL, which is to be expected, as PLP most probably exists as a covalently bound prosthetic group(127,269). One of the substrates for the presumed reverse reaction catalysed by KBL is glycine, and addition of this substrate to KBL caused dramatic decrease in elution volume. The corresponding molecular weight suggests that KBL eluted as an octamer in this experiment. The experiment was repeated (data not shown) and there was a very similar apparent molecular weight of 379.5 kDa. A mixture of TDH and KBL, with NAD^+, L-threonine, PLP and CoA, the final substrate required for catalysis of 2-amino-3-ketobutyrate to acetyl-CoA and glycine, eluted with a single peak that corresponded to a molecular weight of 37.6 kDa. The elution volume suggested that monomers of the two proteins co-eluted as previously.

Looking at the data listed in Table 3.1.4 permits a deeper exploration of the effect of TDH and KBL on each protein's quaternary structure. In the absence of the second protein, TDH and KBL appear to elute in their dimeric forms. Both TDH and KBL appear to have different quaternary structures in the presence of their substrates. The same is true, when the two proteins are tested together in the presence of substrates. When a mixture of TDH and KBL was tested without adding any substrates two peaks were seen on the chromatogram, and SDS-PAGE and silver staining was used to confirm the presence of both TDH and KBL in the fractions corresponding to each peak. Whilst the elution volume at the second peak suggested the presence of TDH and KBL dimers, the earlier peak corresponded to either a multimer of one

protein, or a multi-enzyme complex of the two proteins together. Silver staining suggested that both proteins were present in the sample (see Figure 3.1.32).

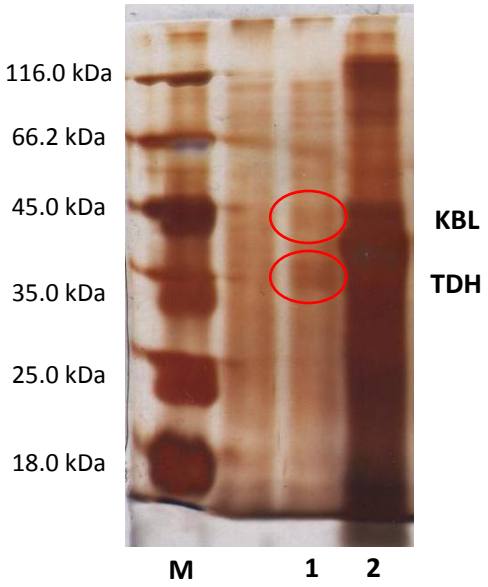


Figure 3.1.32 - silver stain of an SDS-PAGE gel indicating the presence of TDH and KBL in fractions that corresponded to peaks on the chromatogram resulting from the size-exclusion chromatography of a mixture of TDH and KBL. 1 = fraction at 12-14ml elution (peak at 12.5ml). 2 = fraction at 16-18ml(peak at 16.75ml). The protein bands in Lane 1 are light, suggesting that the concentration of both proteins was very low. This is consistent with the fact that the corresponding peak was low. In Lane 2, which analyses a fraction corresponding to a much larger peak, the bands are less distinguishable, due to over-staining in this lane. As the protein concentrations were high, bands corresponding to both proteins can still be seen.

The cross-linking studies carried out using dimethyl suberimidate (DMS) provide more evidence in regards to the oligomeric states of TDH and KBL (see methods in Section 2.1.4.2). In Figure 3.1.33 it can be seen that there are protein bands representing the TDH dimer and the monomeric form. Likewise, there are protein bands representing KBL monomers and dimers on the gel. The concentration of the protein did not appear to affect these results. These findings add strong support to the hypothesis that the native oligomeric states of both TDH and KBL are dimers. There appear to be light protein bands near the top of the gel for which the molecular weight cannot be determined. It is likely that these bands are formed from non-specific cross-linking by DMS.

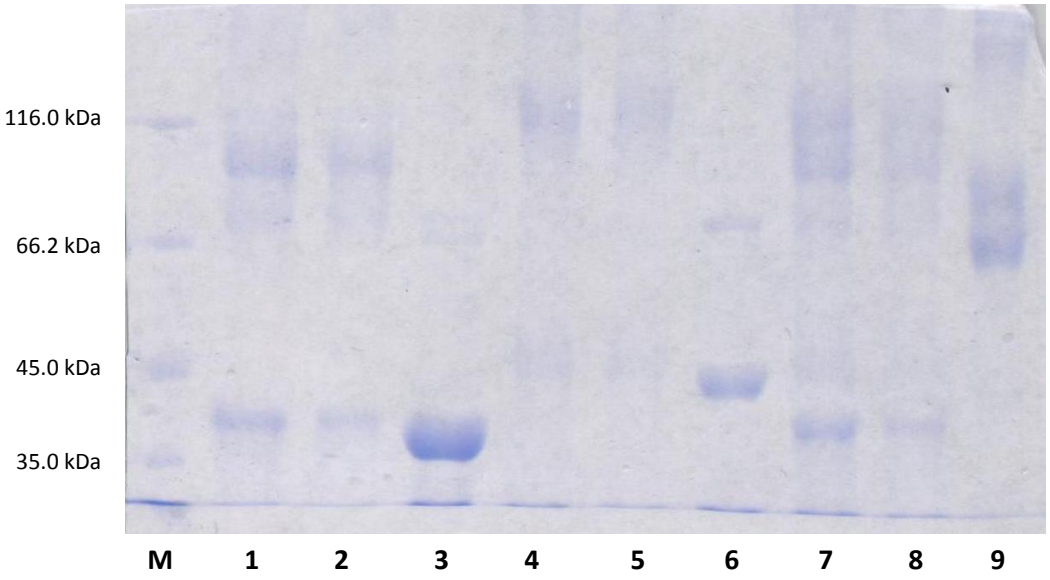


Figure 3.1.33 - SDS PAGE analysis of cross-linking experiments with TDH and KBL, on a 9% polyacrylamide gel. 1 = TDH (2.6 x 10^{-2} mM[1mg/ml]); 2 = TDH (1.3 x 10^{-2} mM [0.5mg/ml]); 3 = TDH (2.6 x 10^{-2} mM [1mg/ml]) non-cross-linked control; 4 = KBL (2.2 x 10^{-2} mM [1mg/ml]); 5 = KBL (1.1 x 10^{-2} mM [0.5mg/ml]); 6 = KBL (2.2 x 10^{-2} mM [1mg/ml]) non-cross-linked control; 7 = TDH (2.6 x 10^{-2} mM [1mg/ml])and KBL (2.2 x 10^{-2} mM [1mg/ml]); 8 = TDH (1.3 x 10^{-2} mM [0.5mg/ml]) and KBL (1.1 x 10^{-2} mM [0.5mg/ml]); 9 = BSA control (1.5 x 10^{-2} mM [1mg/ml]). Note that the protein bands of samples exposed to DMS appear higher due to covalent binding with DMS, leading to a higher molecular weight. This is also the cause of there being more than one band corresponding to a particular oligomer. The results indicate that both TDH and KBL form dimers in solution, but there is no evidence here suggesting that they form a complex together. Dimerisation was not affected by the protein concentration used in these assays.

Experiments conducted with TDH and KBL together do not suggest that the two proteins are complexed in solution, as there are no additional bands present (Figure 3.1.33, lane 7 and 8), compared to when samples of the two proteins alone were tested (Figure 3.1.33, lanes 1, 2, 4 and 5). This conflicts with the finding presented in Figure 3.1.32. Therefore, the existence of a TDH-KBL complex can neither be confirmed nor rejected, based on these findings.

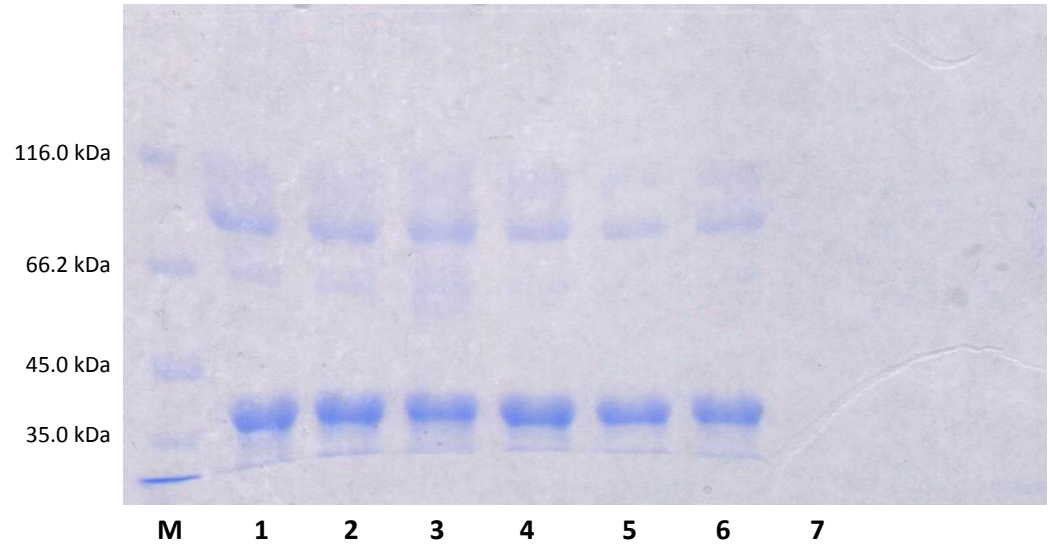


Figure 3.1.34 - SDS PAGE analysis of cross-linking experiments with TDH (at 2.6 x 10^{-2} mM [1mg/ml] concentration) on a 9% polyacrylamide gel. 1 = TDH; 2 = TDH and 1mM NAD^+; 3 = TDH and 10mM NAD^+; 4 = TDH, 1mM NAD^+ and 30mM L-threonine; 5 = TDH, 10mM NAD^+ and 30mM L-threonine; 6 = TDH, 10mM NAD^+ and

15mM L-threonine; 7 = 10mM NAD^+ and 30mM L-threonine control. The oligomerisation of TDH did not appear to be affected by the presence of substrates. All TDH samples showed strong bands representing the dimer and the monomer.

As in the size-exclusion chromatography experiments, the influence of substrates on the oligomeric states of TDH and KBL was investigated. As can be seen in Figure 3.1.34, the oligomerisation of TDH, as indicated through cross-linking with DMS, does not appear to be significantly affected by NAD^+ and/or L-threonine.

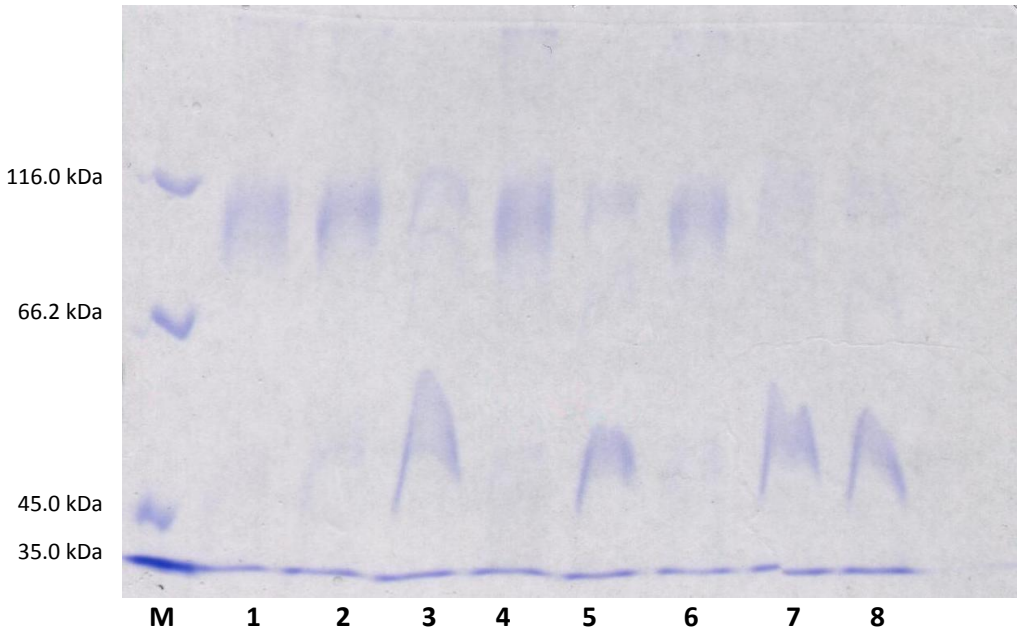


Figure 3.1.35 - SDS PAGE analysis of cross-linking experiments with KBL (2.2 x 10^{-2} mM [1mg/ml]) on a 9% polyacrylamide gel. 1 = KBL; 2 = KBL and 1mM PLP; 3 = KBL and 30mM glycine; 4 = KBL and 5mM A-CoA; 5 = KBL, 1mM PLP and 30mM glycine; 6 = KBL, 1mM PLP and 5mM A-CoA; 7 = KBL, 1mM PLP, 5mM A-CoA and 30mM glycine; 8 = KBL, 5mM A-CoA and 30mM glycine.

Incubation of A-CoA and/or PLP with KBL and DMS does not seem to have an effect on the apparent oligomerisation of the protein. Conversely, glycine appears to either inhibit dimerization or to promote dissociation of KBL. In the absence of glycine, KBL appears predominantly as a dimer. However, in the presence of glycine the majority of KBL is present as the monomer. There is the possibility that glycine interacted with DMS, as the reagent reacts with free amines. However, the reaction buffer contained 200mM of Tris, and this did not inhibit the other cross-linking reactions.

The pull-down assay using His-tagged KBL and native TDH was in agreement with the data from the cross-linking studies and the majority of the data from size-exclusion chromatography studies, that TDH and KBL do not appear to form a complex with each other in solution. When applied to the Ni-NTA column, non-tagged TDH was washed through with the wash buffer. Meanwhile, KBL was retained on the column, and elution with the 250mM imidazole elution

buffer caused KBL to elute alone, demonstrating that the two proteins had not bound to each other or remained bound to each other.

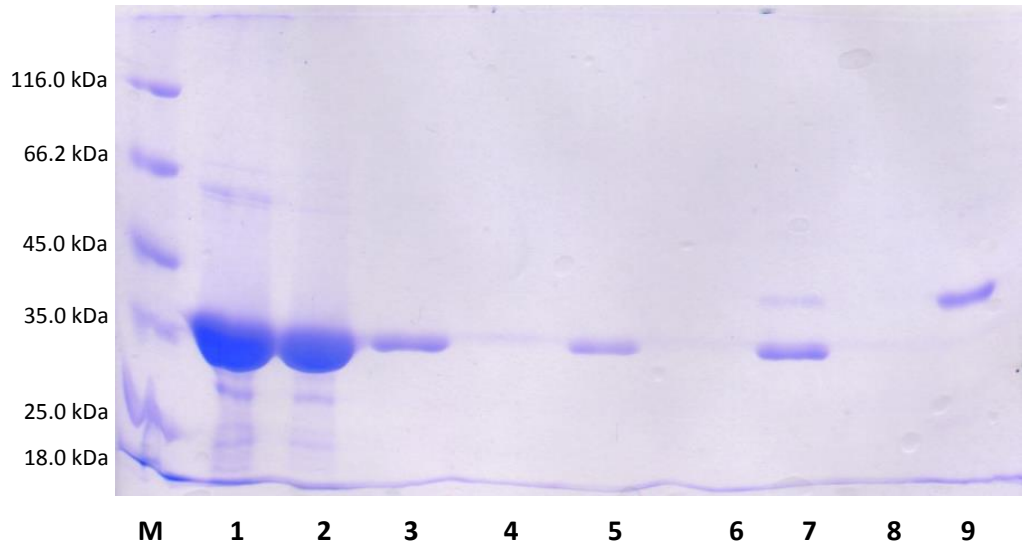


Figure 3.1.36 - SDS PAGE analysis of the pull-down assay to investigate complex formation between TDH and KBL. 1 = original TDH sample; 2 = TDH and thrombin; 3 = TDH removed by washing the solution through a Ni-NTA column and benzamidine column attached end-to-end; 4 = eluate collected on elution of histidine tag from Ni-NTA column; 5 = elution of thrombin (MW approx. 37 kDa) from the benzamidine column; 6 = solution flowing through Ni-NTA column as KBL is loaded; 7 = solution flowing through the Ni-NTA column as TDH is loaded; 8 = solution flowing through the Ni-NTA column as the column is washed with binding or wash buffer; 9 = eluate containing KBL after applying elution buffer to the column. The His-tag was successfully removed from TDH and thrombin was separated from this mixture. His-tagged KBL was retained on a Ni-NTA column, but the untagged TDH was not retained, despite being incubated in the column with KBL overnight. This suggests that a stable complex does not form between the two proteins.

The dimeric nature of TDH and KBL has been confirmed by size-exclusion chromatography, cross-linking and X-ray crystallographic models (for TDH). Although the data on the existence of a TDH-KBL complex are inconclusive, the findings presented here give a strong indication that a number of different oligomeric states exist for both TDH and KBL. Evidence from size-exclusion chromatography indicates that L-threonine promotes the dissociation of TDH, either through binding or through the catalytic action of the enzyme on it. Evidence from size-exclusion chromatography and cross-linking studies also suggest that glycine effects the oligomerisation of KBL. Finally, there is a possibility that TDH and KBL do affect each other's quaternary structure, as indicated by the results of size-exclusion chromatography experiments. The inability to grow complexed crystals of TDH and KBL, adding to the negative cross-linking and pull-down assay results, suggests that a multi-enzyme complex does not exist. However, this does not rule out the existence of a transitory complex, so this cannot be confirmed using these data alone.

3.1.7 Ligand binding and Catalysis

As can be seen in Figure 3.1.22 and Figure 3.1.23, the TDH substrates NAD^+ and L-threonine occupy a deep cleft in the protein. The L-threonine binding pocket and the pocket that is occupied by the nicotinamide group of NAD^+ are partially separated by the side chain of Met81.

Whilst the nicotinamide and the adjacent ribose groups bind in deep pockets of the protein, the rest of the NAD molecule faces the exterior of the protein. NAD binds TDH through a combination of hydrogen bonds and hydrophobic interactions with several main chain and side chain atoms of TDH residues. The interactions with important TDH residues are displayed in Figure 3.1.37 and Figure 3.1.38. In addition to the interactions that NAD makes with the protein, the cofactor also hydrogen bonds with several water molecules. This is particularly the case for the phosphate groups, the ribose hydroxyl groups and the nicotinamide amide.

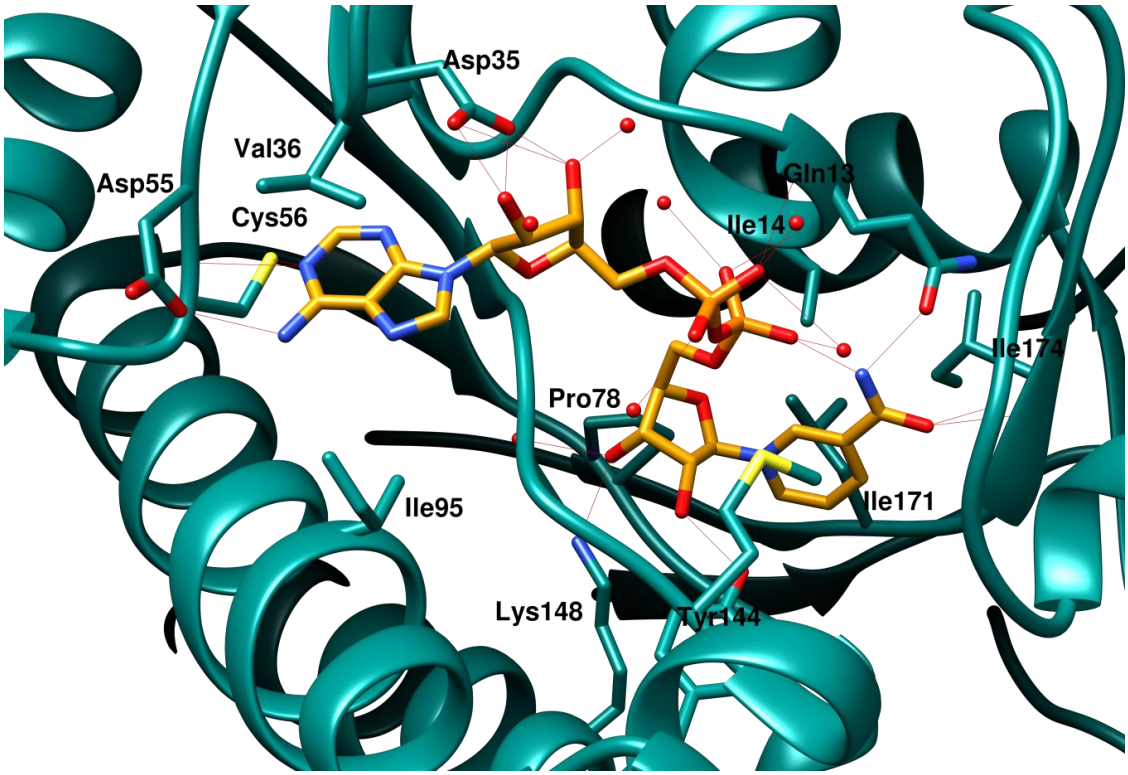


Figure 3.1.37 - Binding of NAD^+ to TDH. The interaction between TDH and NAD involves a combination of hydrogen-bonding and hydrophobic interactions. TDH is shown in a ribbon representation with stick representations of the side-chains of residues that appear to interact with NAD. Water molecules are shown as red spheres at the position of the oxygen atom. Hydrogen bonds are shown by red lines.

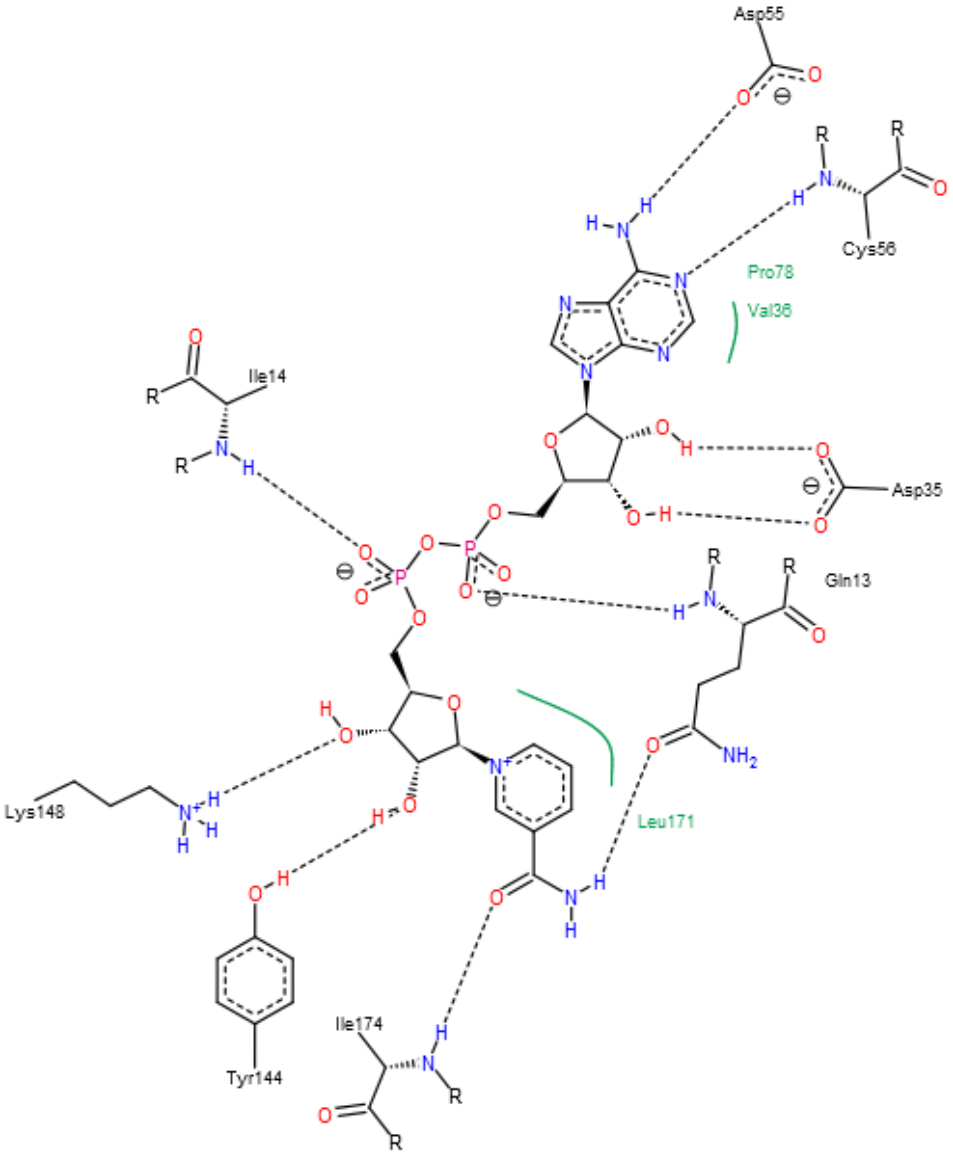


Figure 3.1.38 - binding interactions between TDH residues and NAD^+. Hydrogen bonds are indicated by straight dotted lines. Curved green lines indicate hydrophobic interactions, and the residues involved are also labelled in green text. The diagram was created using PoseView(270–272), and the ligand files were converted to .sdf file format from .pdb files using Open Babel(227).

The X-ray crystallographic model TNADH4, revealed the binding mode of L-threonine. The enzyme was co-crystallised with NADH to prevent turnover. L-threonine is bound predominantly through hydrogen bonds with active site residues. The carboxyl group of L-threonine appears to hydrogen bond with side chain and main chain atoms of Ser82 and Thr186. The amine group of L-threonine is seen to hydrogen bond with the side chain hydroxyl group of Thr86, with the nicotinamide carbonyl of NADH and with a nearby water molecule. The side chain hydroxyl group of L-threonine is hydrogen bonded to Thr119 and Tyr144, whilst the methyl group is directed towards the hydrophobic amino acid Trp280. This aligns the β-carbon with the C4 atom of the nicotinamide ring. The relevance of this will be discussed in more detail below. All

of the residues involved in TDH binding are highly conserved (see section 3.1.3), so they are likely to be essential. Another interesting feature of this model was that the flexible Loop 1 is closed over the active site, meaning that the exterior of the enzyme is no longer accessible to L-threonine. This was the case in both TDH monomers in the asymmetric unit of the crystal model, both of which were bound to NADH and L-threonine.

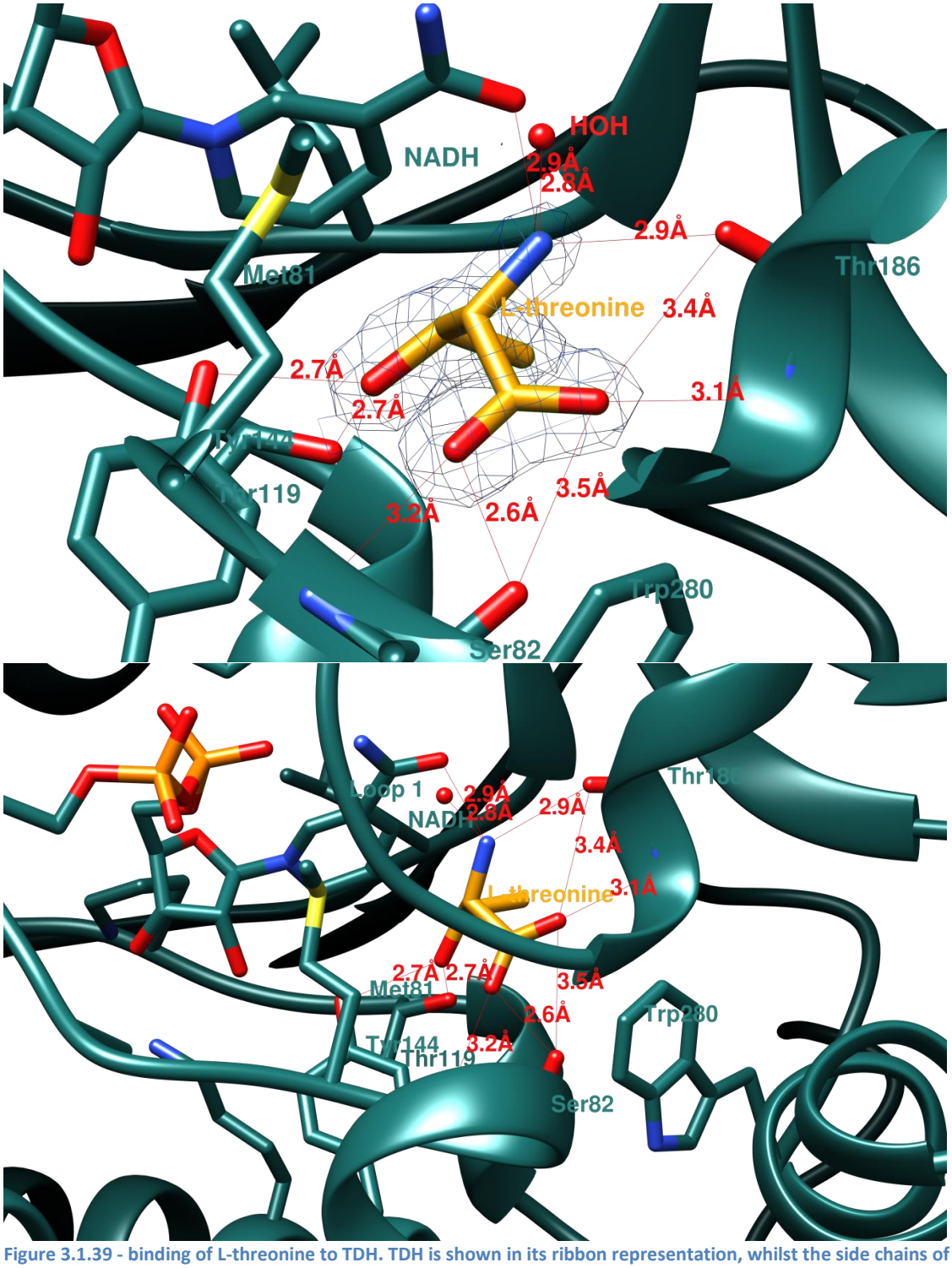


Figure 3.1.39 - binding of L-threonine to TDH. TDH is shown in its ribbon representation, whilst the side chains of interacting amino acids and NADH are represented as sticks. Red lines signify hydrogen bonds, for which the putative bonding distances are labelled in red text. The top image is a close-up that clearly shows the

interactions between L-threonine and different TDH residues. The bottom image shows how Loop 1 is closed over the active site, sealing it off from the external environment of the enzyme.

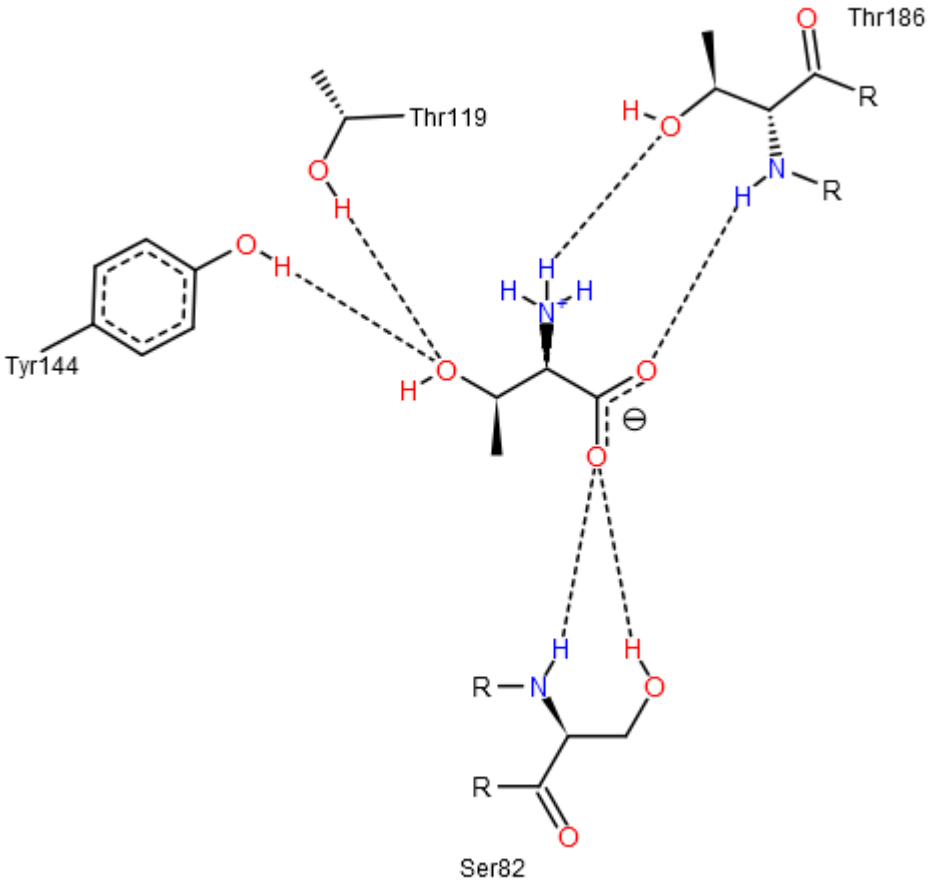


Figure 3.1.40 - interactions between TDH residues and L-threonine, as determined from interatomic distances in the X-ray crystallographic model TNADH4. Straight dotted lines indicate hydrogen bonds. This image was generated by PoseView(270–272) and the ligand file was converted from the .pdb file to .sdf file format using Open Babel(227).

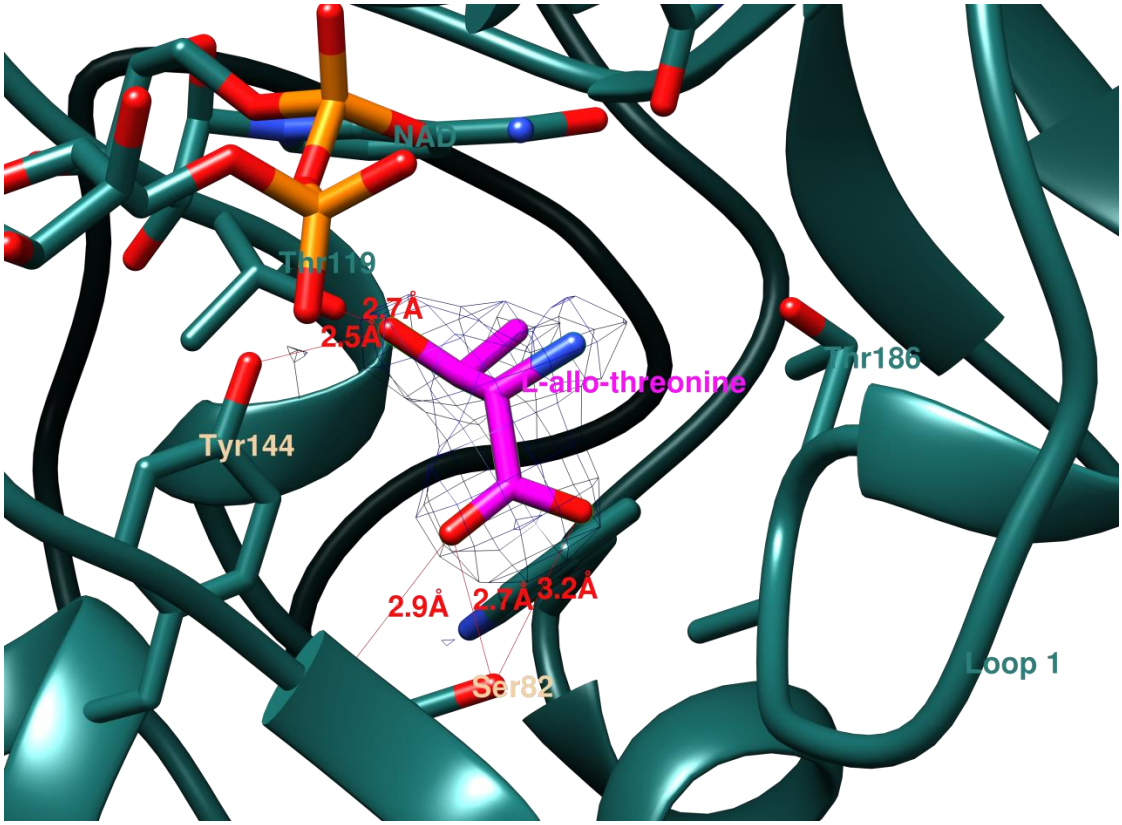


Figure 3.1.41 - binding of L-*allo*-threonine to TDH, observed in X-ray crystallographic model AAT1. TDH is shown in a ribbon representation, and the side chains of interacting amino acids, NAD^+ and L-*allo*-threonine are in stick representations. In contrast to the L-threonine-bound TDH structure, Loop 1 is in its open conformation.

The binding mode of L-*allo*-threonine that was found in the X-ray crystallographic model AAT1 is very similar to that of L-threonine. However, the change in the stereochemistry at the β-carbon means that its orientation is changed slightly. The carboxyl group of L-*allo*-threonine hydrogen bonds with the side chain hydroxyl group and the main chain nitrogen atom of Ser82 in a similar manner to L-threonine. It also seems likely that this group interacts with Thr86. The side-chain hydroxyl group of L-*allo*-threonine is hydrogen-bonded by Thr119 and Tyr144, but the *S* configuration of this group means that the distance between the β-carbon and the C4 of the NAD nicotinamide group is increased. The difference in stereochemistry of L-*allo*-threonine (*2S,3S* configuration), compared to L-threonine (*2S,3R* configuration), has been cited as a reason for the lack of catalysis of L-*allo*-threonine by TDH(200). The difference in stereochemistry also affects the entire orientation of the molecule. As a result, the amino group of L-*allo*-threonine is at a further distance from Thr186 and the nicotinamide amide than the amine group of L-threonine. In model AAT1, the flexible loop 1 is in an open conformation in a monomer where L-*allo*-threonine is bound, which contrasts with the finding that the flexible loop is closed where L-threonine is bound in NADH4.

In another TDH structure model, which was solved in this laboratory by Dr Peter Erskine, pyruvate is also seen to occupy the L-threonine binding site. It seems to exploit similar binding

partners to L-threonine: Ser82, Thr119, Tyr144 and Thr186. As was observed with L-threonine-bound structures, Loop 1 was in its closed conformation.

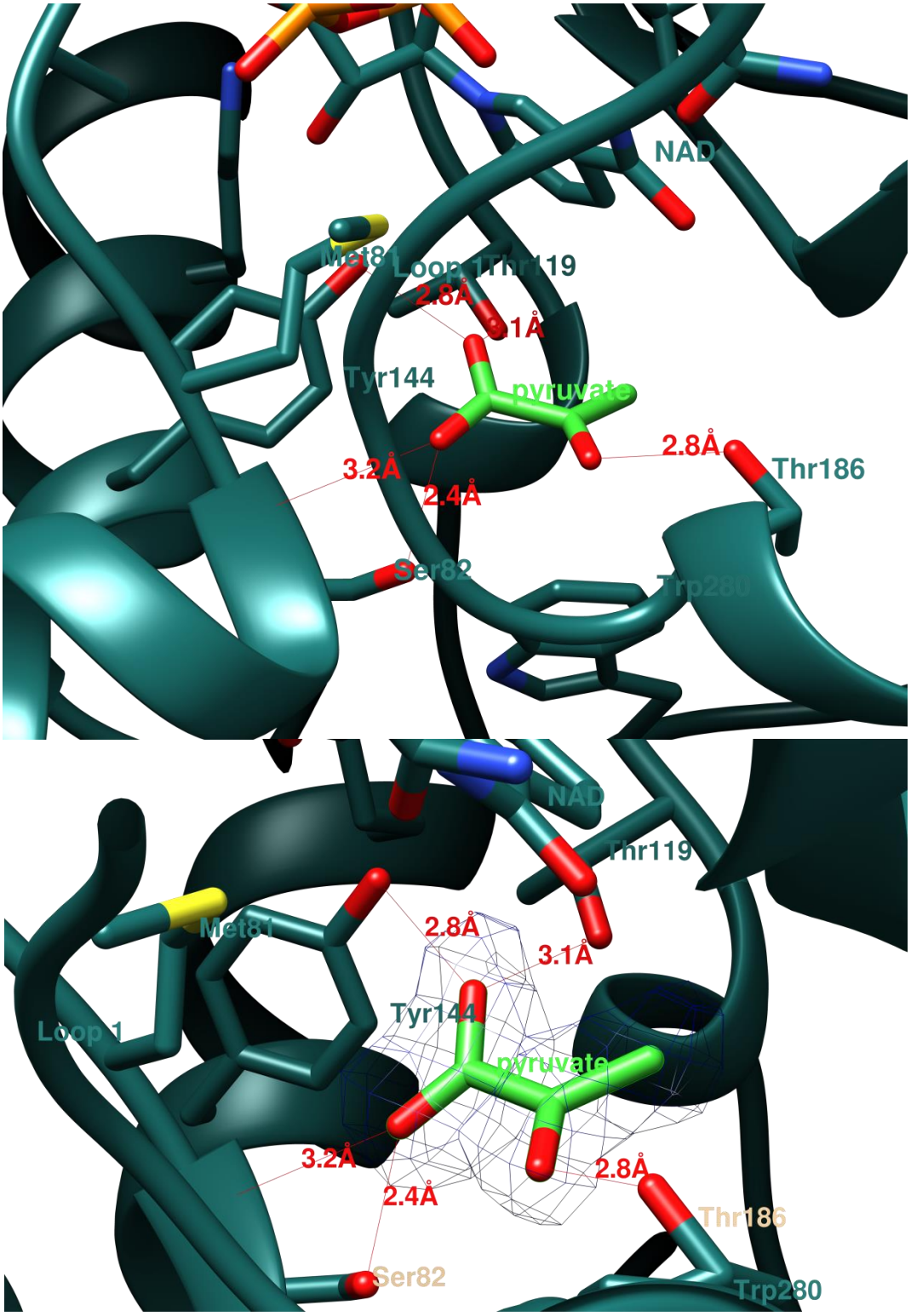


Figure 3.1.42 - binding of pyruvate to TDH, as observed in an X-ray crystallographic structure. As can be seen in the upper image, Loop 1 is in its closed conformation when pyruvate is bound. The lower image shows the binding interactions from a closer perspective. Pyruvate shares some of the same binding partners as TDH.

3.1.7.1 Mechanism of Action

Having insight into the three-dimensional structure of TDH bound to its substrates provides information alluding to the mechanism of catalysis. The nicotinamide and L-threonine binding pockets lie adjacent to each other and, when occupied by the respective substrates, they allow NAD and L-threonine to be in close enough proximity for catalysis. The putative mechanism of reaction of TDH involves transfer of a hydride from the β-carbon of L-threonine to the C4 atom of the nicotinamide ring. The data presented here confirm that the atoms involved in this reaction are well aligned and at a distance that would permit such a reaction to happen (see Figure 3.1.22 and Figure 3.1.39). The catalytic process involves the extraction of a proton from the side chain hydroxyl group of L-threonine, which is converted to a carbonyl group as the negatively-charged oxygen atom forms a double bond with the electron-deficient β-carbon. The removal of the proton from the hydroxyl group requires a number of structural features to be present in the enzyme, and a number of said features are found to be conserved in SDR enzymes. One such feature is the presence of a "catalytic triad", which in TDH from *T. brucei* includes Thr119, Tyr144 and Lys148. The importance of these residues has been demonstrated above. Another feature is a chain of water molecules that is believed to form a proton relay that transfers the proton to the bulk solution, allowing the enzyme to catalyse another reaction. This water chain requires a certain orientation of residues to bind the water molecules. One widely conserved feature in SDRs is an interaction between an asparagine side chain and the main chain of another residue. In this case, the interaction occurs between Asn96 and Ile80. The binding of the Asn96 causes a kink in the α-helix that it belongs to and changes the orientation of its carbonyl group. This allows the carbonyl group to hydrogen bond one of the water molecules in the water chain. Figure 3.1.43 shows that this feature was also confirmed to be present in TDH from *T. brucei*.

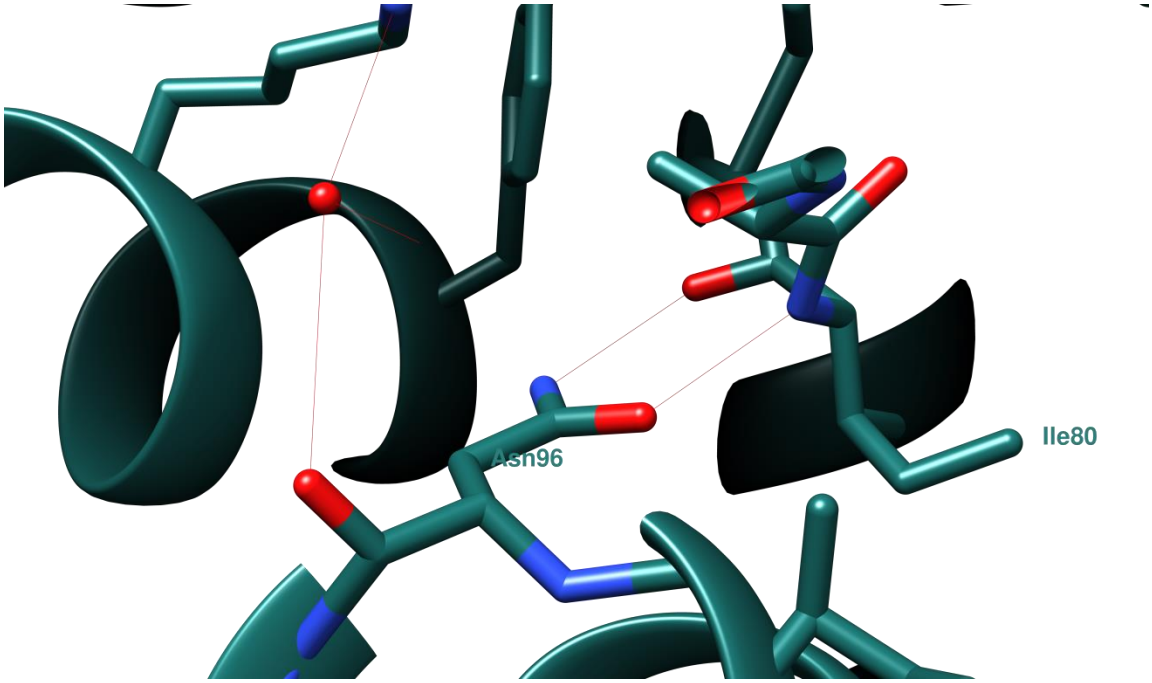


Figure 3.1.43 - hydrogen bonding between the side chain of Asn96 and main chain atoms of Ile80 changes the orientation of the carbonyl of Asn96, allowing it to hydrogen bond a water atom that is proposed to be important in the proton relay of SDR enzymes.

In the putative proton relay, the proton is extracted from L-threonine by Tyr144. The proton then passes to a ribose oxygen on NAD, then to the amine side-chain of Lys148 and then on to a chain of water molecules, including the water molecule that is hydrogen bonded by the main-chain carbonyl of Asn96. A full water chain that would allow the proton to be transferred to the surrounding solution was not identified in any of the structures presented here. However, Figure 3.1.44 shows how the existence of a proton relay system is highly likely. The distances between the important amino acids and the water molecules seem to confirm this. Part of a water chain was present in each structure model solved.

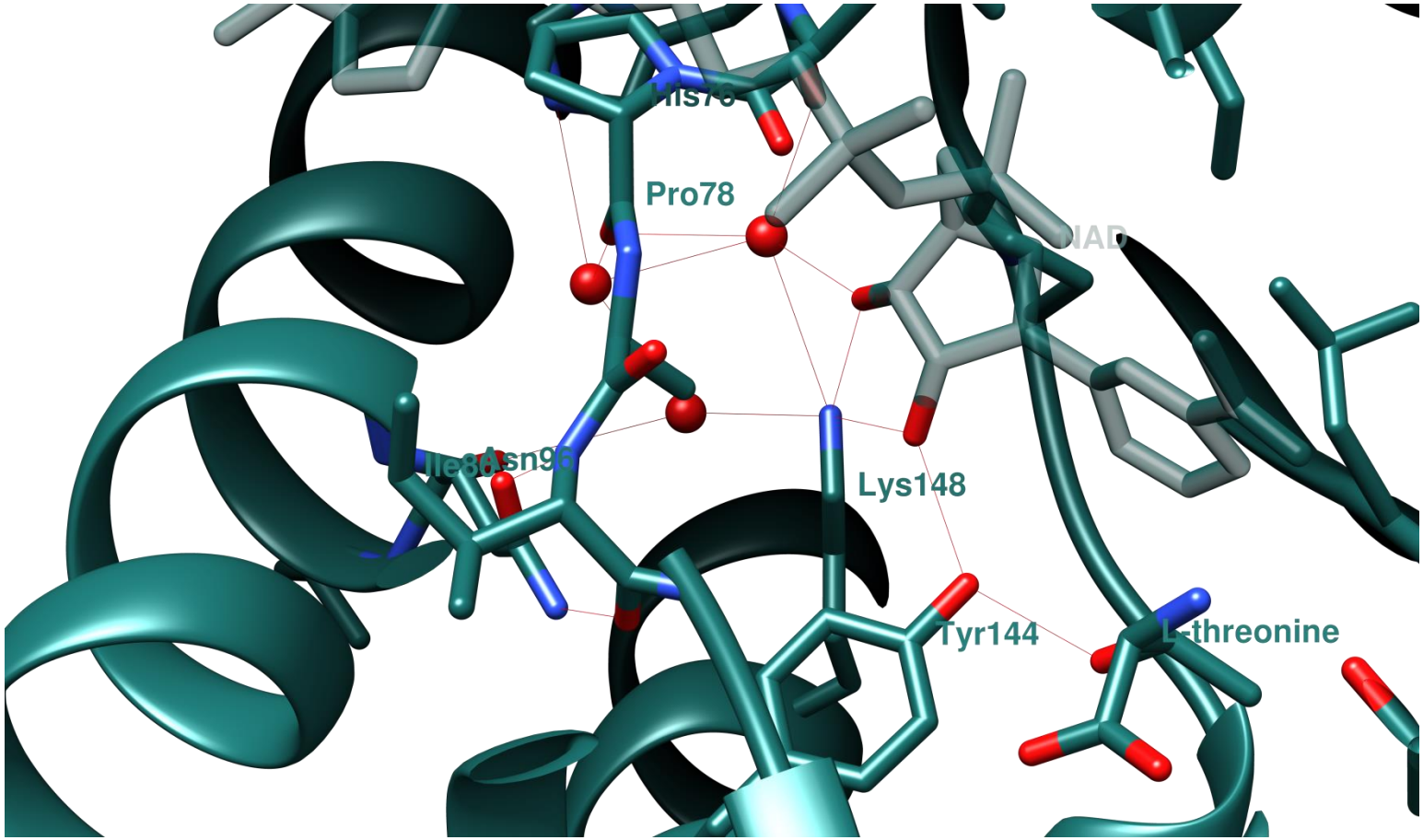


Figure 3.1.44 - atoms within hydrogen-bonding distance of each other, which may form part of a proton relay system. A proton extracted from the hydroxyl group of L-threonine could pass from Tyr144, to the NAD ribose oxygen, then to Lys148, to the water molecule that is hydrogen-bonded to Asn96, and then onto other nearby water molecules. A complete water chain was not observed and it was not clear if a direct network of water molecules linking the Lys148 to the bulk solvent was present.

3.1.7.2 *Other features observed in crystallographic models*

Molecules from the storage buffer and from the crystallisation solutions co-crystallised with TDH were found to be bound to it. Glycerol, acetate, sulphate ions and sodium ions were all observed in TDH structure models. The molecules occupied various positions, but a few positions were occupied in a repeatable manner. Most notably, glycerol and acetate bound within the L-threonine binding site (see Figure 3.1.45). Glycerol was also found to bind to the same regions in different TDH models (shown in Figure 3.1.46).

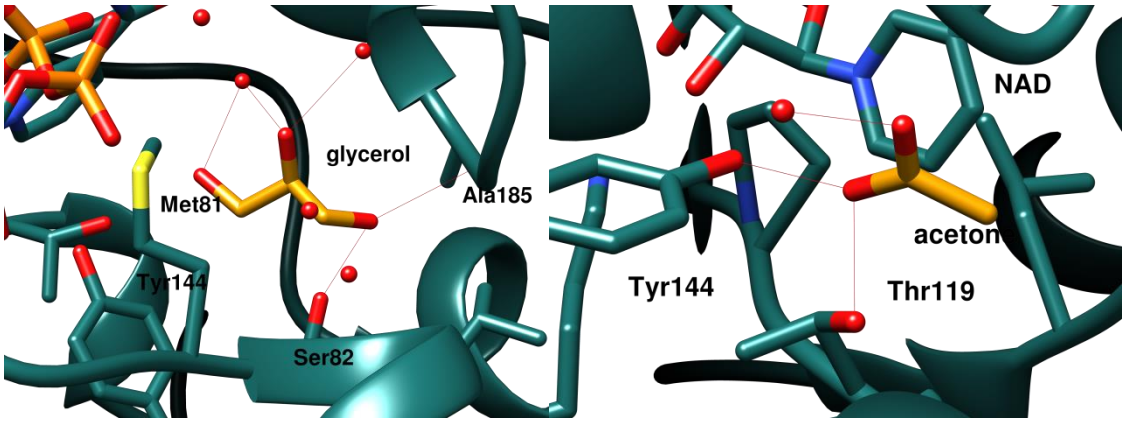


Figure 3.1.45 - binding positions in the active site of TDH occupied by glycerol (left) and acetone (right). These positions were observed in more than one monomer and/or structure model. The molecules interact with the same residues as L-threonine. Glycerol binds predominantly with Ser82, whilst acetone binds with Thr119 and Tyr144.

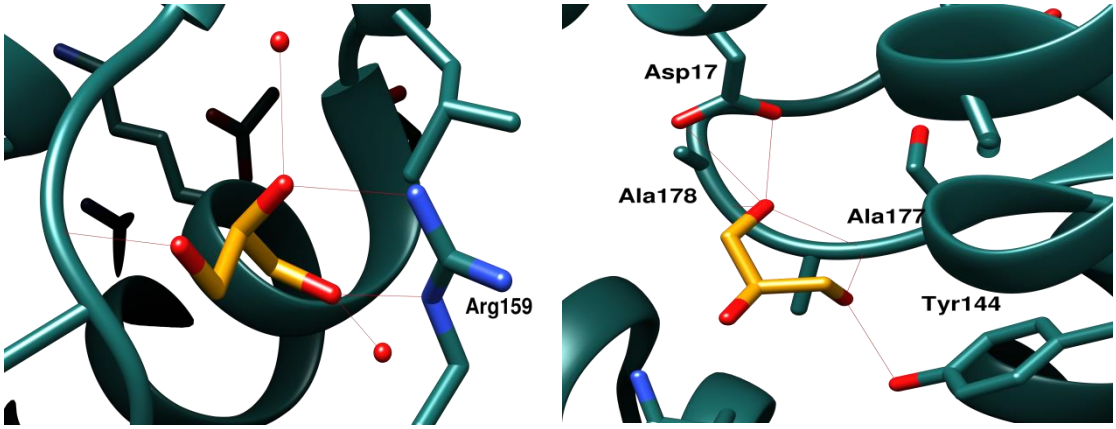


Figure 3.1.46 - two more binding positions of glycerol on TDH. Glycerol molecules were often found adjacent to the residue Arg159 (shown left) and Asp17 (shown right).

Typically, TDH monomers that were a part of different dimers would be separated from other monomers by an ordered array of water molecules. However, one interaction between two monomers in separate dimers was observed in model AAT1. Interestingly, the interaction is formed between residues of the flexible Loop 2 region (see Figure 3.1.15). The interaction does not appear to be strong enough to be an important physiological interaction, so it is likely to be a result of the crystallisation process.

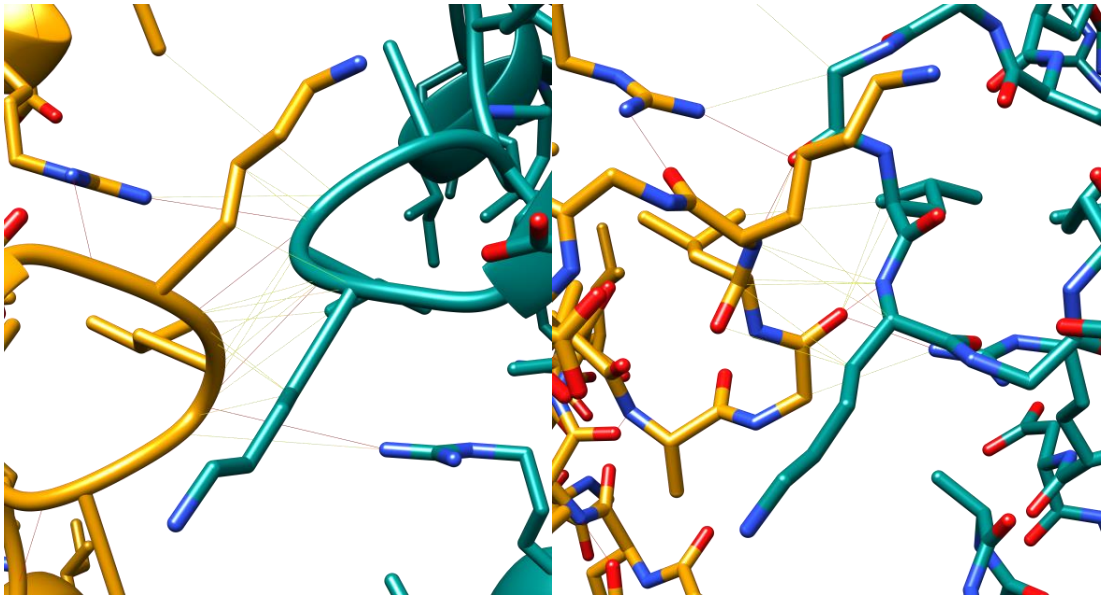


Figure 3.1.47 - an interaction between two TDH monomers at the Loop 2 region (Pro44-Gly50). A ribbon representation with stick-represented side chains (left) and a stick representation (right) are shown.

3.2 TDH kinetics

3.2.1 Kinetic Characteristics of TDH

Initial studies of TDH kinetics in the presence of different combinations of NAD^+ and L-threonine concentrations established the saturating concentration as 10mM and 30mM for NAD^+ and L-threonine, respectively (see methods in Section 2.2.1).

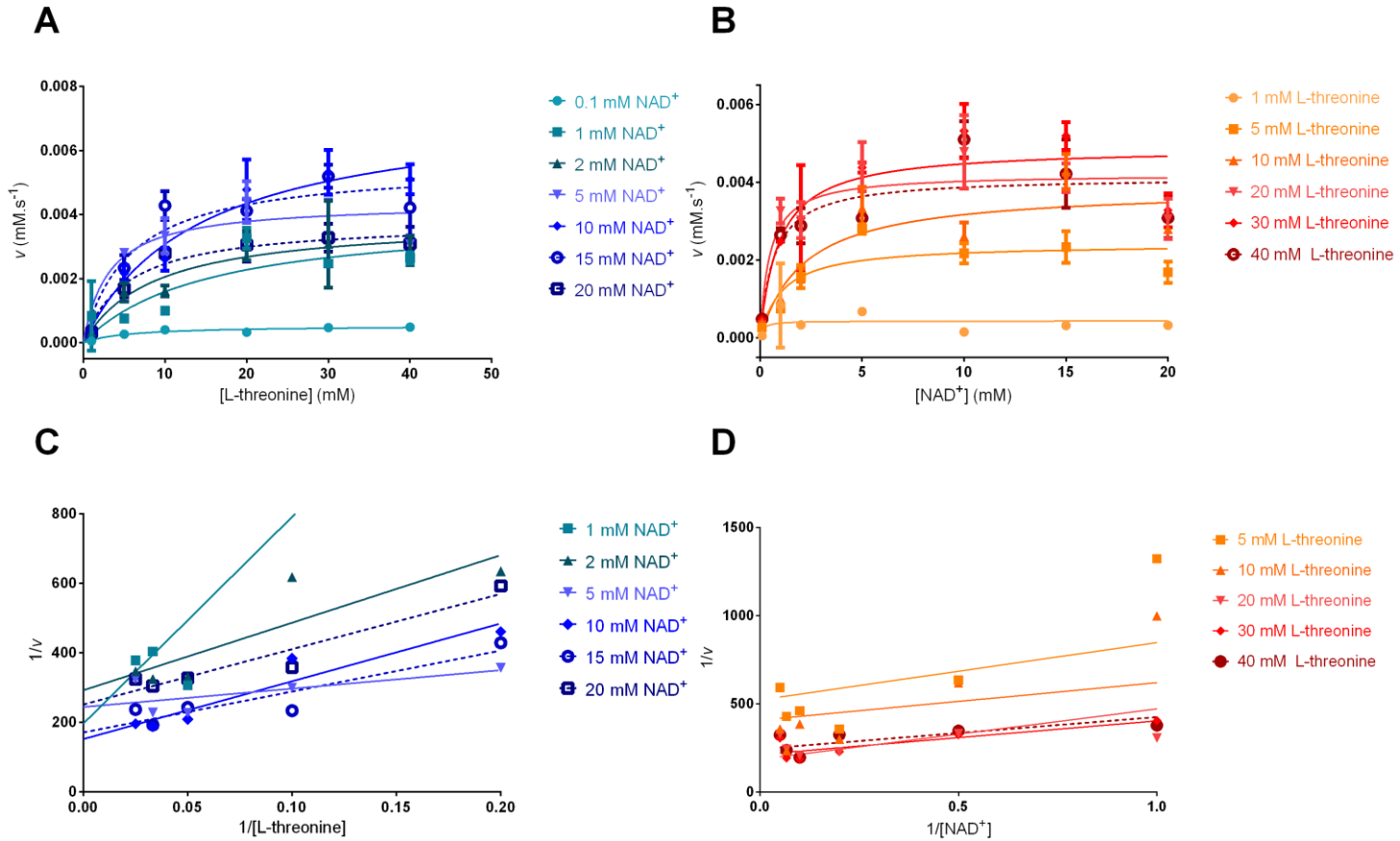


Figure 3.2.1 - kinetic plots and Lineweaver-Burk plots of rate of TDH activity (*v*) as a function of substrate concentration. A - *v* as a function of L-threonine concentration at fixed concentrations of NAD^+; B - *v* as a function of NAD^+ concentration at different fixed concentrations of L-threonine; C - Lineweaver-Burk plot of A; D - Lineweaver-Burk plot of B. Lines of best fit in A and B were fit to the Michaelis-Menten equation and/or Hill equation. Rates of activity generally increased at higher substrate concentrations. Fitted lines or curves are dotted where higher substrate concentrations cause a decrease in activity, relative to a lower concentration. See Figure 2.2.5 for the method of derivation of these graphs.

The patterns of lines in the double-reciprocal plots (Lineweaver-Burk plots) in Figure 3.2.1 indicate that TDH acts by an ordered sequential Bi Bi mechanism of action. This means that one substrate binds to the free enzyme, prior to the binding of the second substrate. The order of binding can be compulsory or random. However, taking into consideration the data from X-ray crystallography, it can be concluded that the observed kinetic behaviour is probably indicative of a compulsory ordered mechanism, with NAD^+ binding first, as illustrated in Figure 3.2.2. The pattern of lines of decreasing slopes (from 1 – 5mM NAD^+) in Figure 3.2.1C suggests that L-threonine binding is dependent on the NAD^+ concentration, and thus, NAD^+ binding to TDH. Conversely, in Figure 3.2.1D, the lines of best fit are more or less parallel with each other, suggesting that the L-threonine concentration, and therefore L-threonine binding to TDH, does not have a large influence on NAD^+ binding. This indicates that NAD^+ is the substrate that must bind to TDH first, making this a compulsory ordered sequential mechanism. The structural data on TDH and its interaction with its substrates, and the data from inhibitory assays (see below) provide further confirmation of a compulsory ordered mechanism of action.

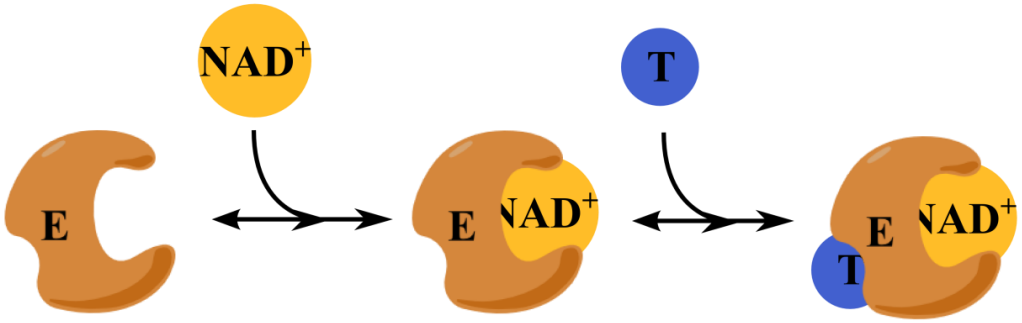


Figure 3.2.2 - Drawing illustrating a Compulsory Ordered Bi Bi mechanism of enzyme catalysis. The first substrate, NAD^+ binds to the enzyme (E), then the second substrate, L-threonine (T), binds.

In Figure 3.2.1, it can be seen that, above certain NAD^+ and L-threonine concentrations, the catalytic rate decreases (see dotted curves and lines). This is caused by inhibition of TDH at high substrate, and subsequently, at high product concentrations. Figure 3.2.3 shows the effects of higher substrate concentrations on TDH activity. Above 10mM NAD^+, the rate of activity decreases significantly. This effect is much more pronounced than the decrease in activity at L-threonine concentrations above 30mM.

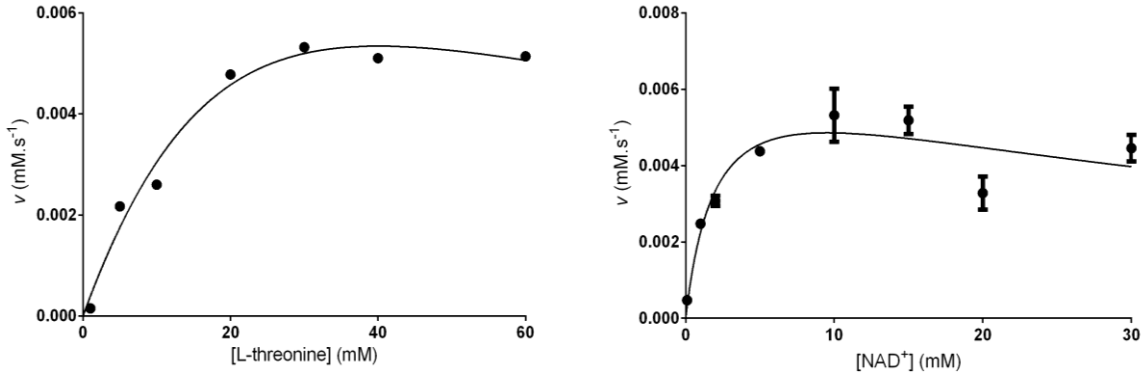


Figure 3.2.3 - Plots of catalytic rate (*v*) as a function of substrate concentration, showing signs of substrate inhibition. Curves are fitted to the data using an equation that accounts for the effect of substrate/product inhibition on rate.

In the case of NAD^+, it is possible that the inhibition is caused by inhibition with the product of the reaction, NADH. The fact that the oxidised and reduced forms of the cofactor are almost identical means that there is a high possibility that TDH has a similar affinity for both. Therefore, as the rate of NADH formation increases above a certain threshold, it begins to inhibit the binding of NAD^+. The data for NAD^+ fit to an equation describing substrate inhibition with an R^2 value of 0.8457. The predicted K_i was 43.5mM and the K_M for NAD^+ was 2.1mM.

Up to concentrations of 10mM, the relationship between NAD^+ concentration and rate conformed to normal Michaelis-Menten kinetics. When fit to the Hill modification of the Michaelis-Menten equation (Equation 11), the calculated Hill coefficient was 0.97 at 37°C and 1.03 at 23°C, suggesting that NAD^+ binding is not cooperative. Slightly different values of K_M, 0.99mM and 0.66mM, were calculated at 37°C and 23°C, respectively. These values are similar to each other, but lower than that calculated by the substrate inhibition equation because they do not account for the inhibitory effect of NADH on the V_{max} that would be achieved in the absence of inhibition.

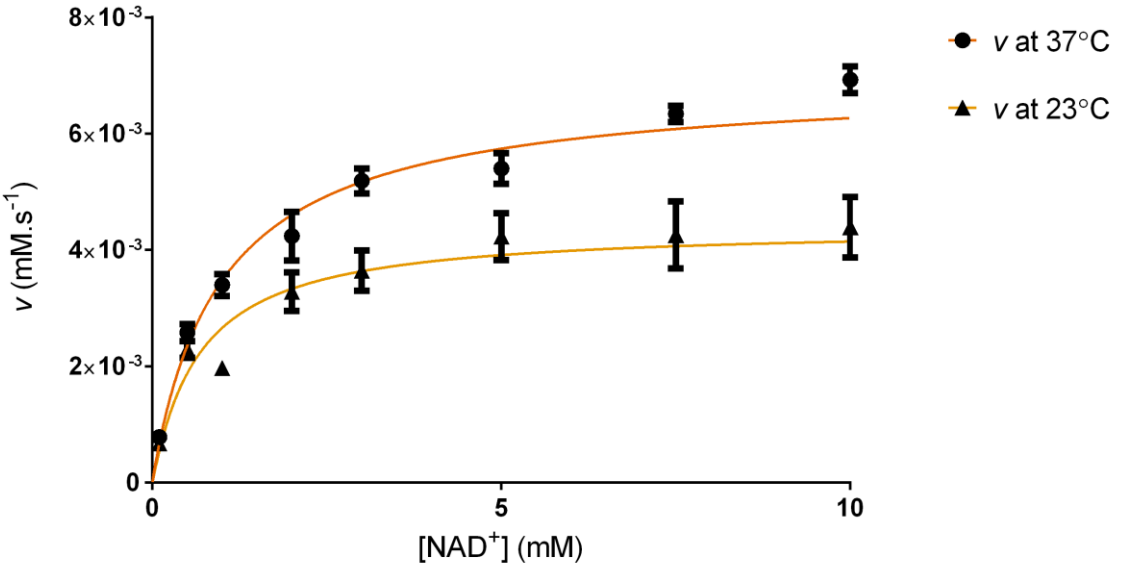


Figure 3.2.4 - plots of *v* against NAD^+ concentration, measured at 37°C and 23°C. The data were fit to equation 11.

In contrast to the data obtained for NAD^+, the data showing the relationship between L-threonine concentration and rate exhibited an atypical, double-sigmoidal pattern that could not be described by the Michaelis-Menten equation or the Hill equation. As the L-threonine concentration approaches 15mM, the rate of activity plateaus. However, as the concentration is increased further, there is a sharp increase in activity up to 30mM L-threonine.

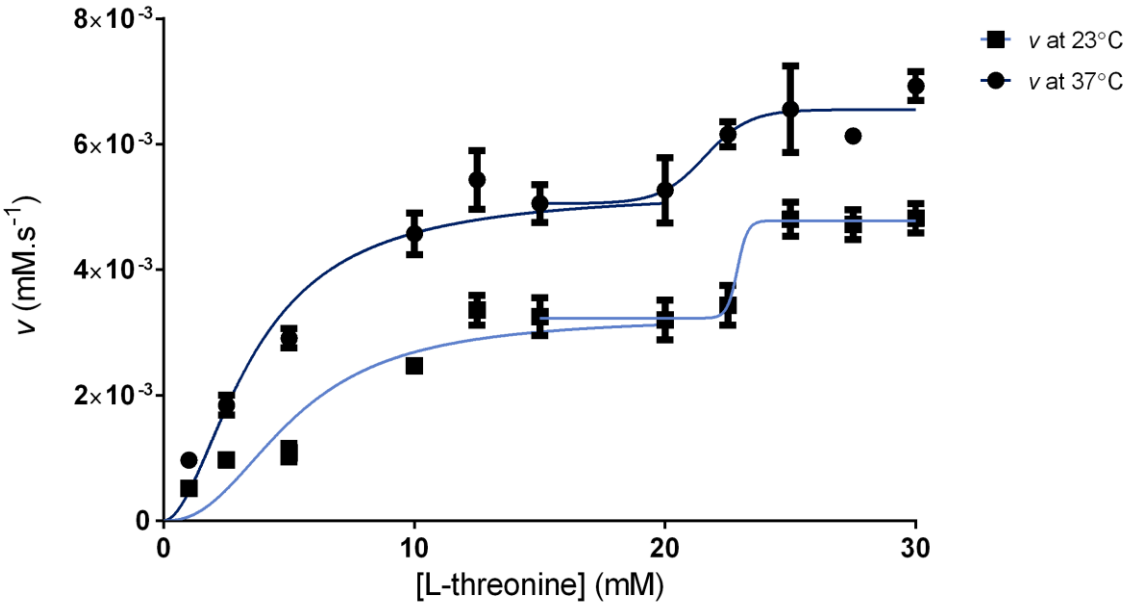


Figure 3.2.5 - plots of *v* versus L-threonine concentration at 23°C and 37°C. Data corresponding to lower and higher L-threonine concentrations were fit separately to equation 11, as the data were not well fit to a single curve.

Using the data analysis software, Origin Pro, the two sigmoidal portions of the curve were fit separately, giving two values each for K_M, V_{max}, the maximal rate of activity at each plateau, and k_{cat}, the turnover number or number of reactions catalysed by each active site. At 37°C: K_M^1 = 5.34mM, V_{max}^1 = 5.22 x 10^{-3}mM.s^{-1}, k_{cat}^1 =14.95s^{-1}; K_M^2 = 21.55mM, V_{max}^2 = 6.56 x 10^{-3}mM.s^{-1}, k_{cat}^2 =18.78s^{-1}. At 23°C: K_M^1 = 8.16mM, V_{max}^1 = 3.39 x 10^{-3}mM.s^{-1}, k_{cat}^1 =9.70s^{-1}; K_M^2 = 22.73mM, V_{max}^2 = 4.78 x 10^{-3}mM.s^{-1}, k_{cat}^2 =13.70s^{-1}. Both curves show evidence of positive cooperativity, with similar Hill coefficients of 4 at 37°C and 4.7 at 23°C being calculated for the first portion of the curve. However, the Hill coefficients calculated for the second portion of each curve are largely different: 23.3 at 37°C and 187.22 at 23°C. The variance in values of *h*, coupled with the unfeasibility of these values as reflections of the number of cooperative active sites, clearly shows that the Hill equation is not entirely valid in this case.

The variable exponent, q, proposed by Kurganov and colleagues(157) can be used to describe the cooperative behaviour of TDH at different concentrations of L-threonine:

$$v = V_{max}\frac{[S]^q}{[S]_{0.5}^q + [S]^q} \qquad \textbf{23}$$

In this description of cooperative enzyme activity, the rate, v, is dependent on the relationship between V_{max}, the maximal rate of activity, [S], the substrate concentration, and the invariable exponent, *q*. $[S]_{0.5}$ is similar to the Michaelis-Menten constant K_M and represents the substrate concentration that produces 50% of the V_{max}. Values of *q* at different concentrations of L-threonine were determined iteratively using data collected at 37°C. As *q* describes the relationship between substrate concentration and cooperativity, the values should be similar at any temperature. When these q values were used to calculate estimates of *v* at 23°C using the modified Hill equation, an R^2 of 0.98 was calculated.

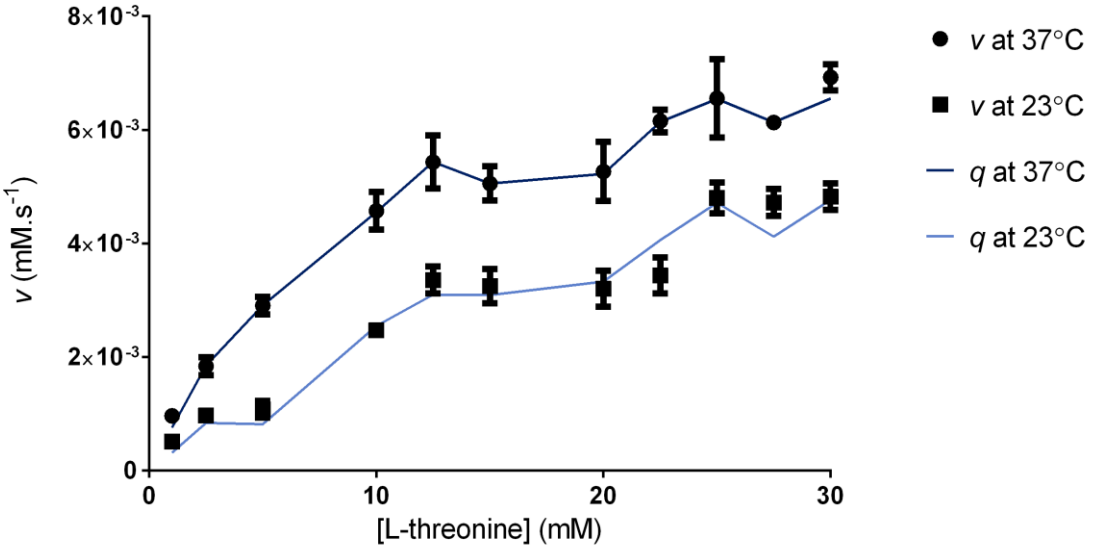


Figure 3.2.6 - plots of *v* versus [L-threonine], with data points predicted using the inconstant exponent *q* (dark and light blue lines). Values for *q* were calculated at each L-threonine concentration from the data collected at 37°C and used to calculate data points, which are represented by a line.

Although *q* does not provide information on mechanisms of cooperativity, it can describe the progressive change in binding affinity, and consequently catalytic activity, and the substrate concentration. Figure 3.2.7 shows a plot of the *q* values corresponding to each L-threonine concentration in the equation 23, which describes the double-sigmoidal curves in Figure 3.2.6.

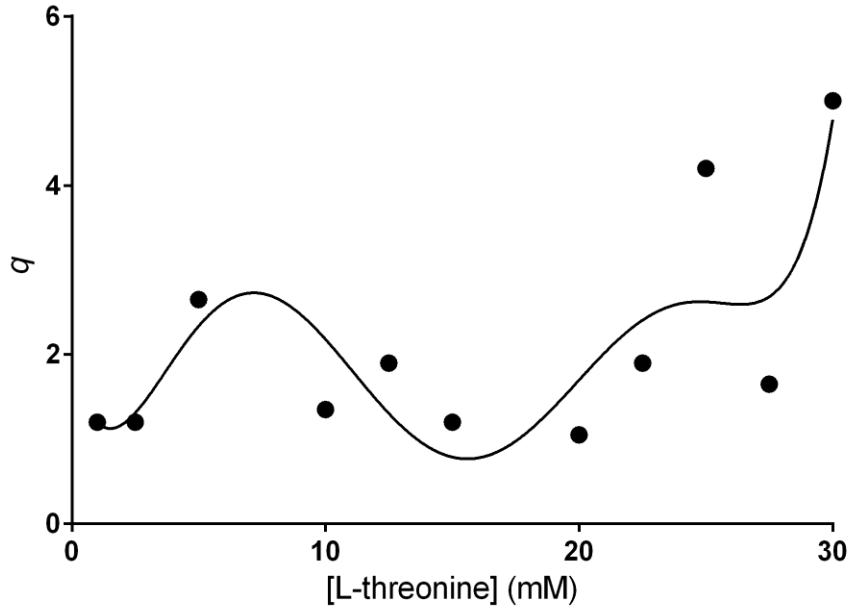


Figure 3.2.7 - plot of *q* against corresponding L-threonine concentrations. A curve described by a polynomial equation was fit to the data to highlight the relationship between *q* and increasing L-threonine concentrations.

In Figure 3.2.7 it can be seen that *q* increases with L-threonine concentration up to 7.5mM, at which point it starts to decrease. The value of *q* reaches a minimum between 15 and 20mM L-threonine, which corresponds to a plateau in activity, before increasing sharply as the concentration is increased further.

An equation to which a double sigmoidal curve can be fitted was proposed by Selwood *et al.* to describe the catalytic activity of human porphobilinogen synthase(273). This equation actually describes a model where the rate of activity is the sum of activity produced by two active sites of different affinity.

$$v = \frac{V_{max1} \cdot [S]}{K_{M1} + [S]} + \frac{V_{max2} \cdot [S]}{K_{M2} + [S]} \qquad \textbf{24}$$

In the equation above, which describes activity of two different forms of an enzyme, K_{M1} and V_{max1}, and K_{M2} and V_{max2} are the half-saturating substrate concentrations and maximal rates of reactions for a high-affinity form of the enzyme and a low-affinity form, respectively. The high affinity form is predominantly responsible for the activity at lower substrate concentrations, whilst the low affinity form has a higher K_M and is responsible for the activity at higher substrate concentrations. As discussed above, when described separately, both halves of the curve show evidence of cooperativity. For this reason, the Hill coefficient *h* was applied to the equation above, so that it resembled a summation of two Hill equations.

$$v = \frac{V_{max1} \cdot [S]^{h1}}{{K_{M1}}^{h1} + [S]^{h1}} + \frac{V_{max2} \cdot [S]^{h2}}{{K_{M2}}^{h2} + [S]^{h2}} \qquad \textbf{25}$$

Two separate Hill coefficients are designated for each form of the enzyme, and hence each part of the curve: *h1* and *h2*. A plot of rate versus L-threonine concentration at 23°C, fit to equation 25 is shown in Figure 3.2.8.

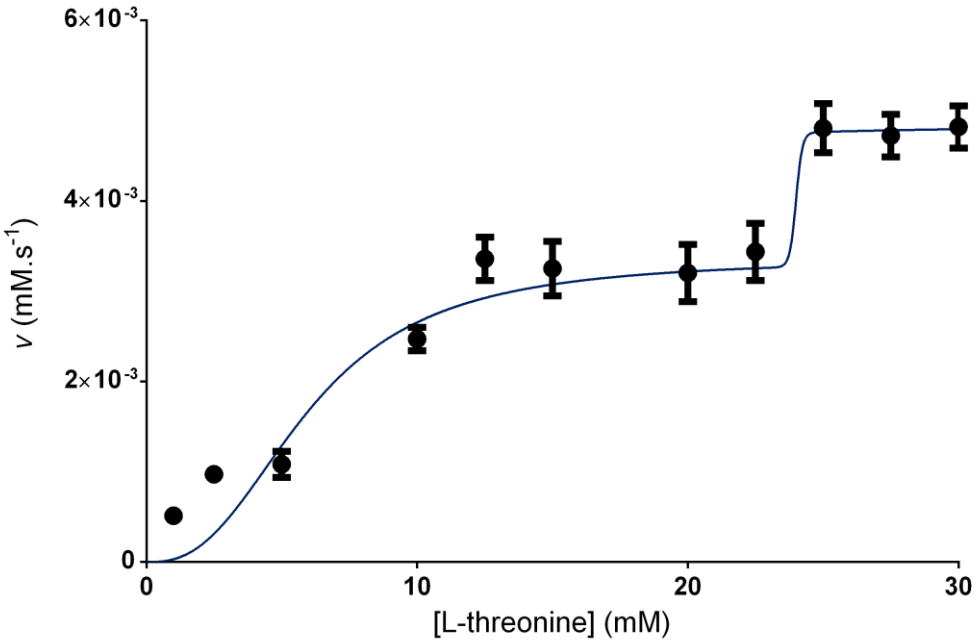


Figure 3.2.8 - plot of *v* versus L-threonine concentration. The data were fit to a curve based on equation 25. R^2 = 0.94. The curve was fit by constraining K_{M1} to 6mM, V_{max1} to 3.4 x 10^{-3}mM.s^{-1}, *h1* to ≥ 2 and K_{M2} to 24mM.

With a R^2 of 0.94, the equation clearly described the data well. However, in order for the data analysis software Prism to fit it, it was necessary to constrain several of the equation's parameters. In particular, it was necessary to constrain $V_{max}{}^1$ to one of the values from the first plateau in the curve, and the two K_M values were constrained using values obtained by separately fitting each half of the curve to the Hill equation, as described earlier.

The results of a test of the ability of different amino acids (at 60mM) to inhibit TDH activity, or to be catalysed by TDH (see Section 2.2.4) are displayed in Table 3.2.1.

Table 3.2.1 - TDH activity in the presence of 60mM of various amino acids. The 'Inhibition' column shows the TDH activity (as a percentage of activity in control conditions) achieved with saturating substrate concentrations and the amino acid. The 'Catabolism' column lists the percentage activity of TDH in the presence of 10mM NAD^+ and 60mM of amino acid.

	Inhibition	Catabolism
Amino acid	Percentage activity ± SD	Percentage activity ± SD
L-alanine	70.76% ± 3.57%	0.11% ± 0.56%
L-cysteine	39.39% ± 4.78%	-0.25% ± 0.80%
L-aspartic acid	68.71% ± 4.20%	1.16% ± 0.52%
L-glutamic acid	74.78% ± 0.08%	0.97% ± 0.51%
L-phenylalanine	86.39% ± 2.32%	0.77% ± 0.84%
glycine	80.44% ± 2.75%	-0.48% ± 0.97%
L-histamine	77.74% ± 0.34%	1.63% ± 0.12%
L-isoleucine	88.30% ± 7.37%	1.82% ± 0.33%
L-lysine	97.69% ± 2.20%	1.67% ± 0.54%
L-leucine	93.14% ± 4.46%	1.75% ± 0.49%
L-methionine	89.34% ± 22.19%	-0.30% ± 2.07%
L-asparagine	91.99% ± 11.85%	-1.31% ± 1.31%
L-proline	96.63% ± 3.73%	-0.35% ± 0.52%
L-glutamine	126.87% ± 2.95%	0.21% ± 0.15%
L-arginine	101.64% ± 5.99%	0.30% ± 0.42%
L-serine	99.22% ± 12.92%	0.91% ± 0.10%
L-valine	92.72% ± 0.02%	-0.27% ± 1.18%
L-tryptophan	99.25% ± 41.58%	1.39% ± 3.17%

Based on the results above, the following amino acids appear to inhibit TDH: L-alanine, L-cysteine, L-aspartic acid, L-glutamic acid, L-phenylalanine, glycine, L-histamine, L-isoleucine and L-leucine. L-cysteine is the most significant inhibitor of this list, causing more than 60% inhibition (39.4% residual activity). Studies on the ability of TDH to catalyse an oxidative reaction in the absence of L-threonine, but in the presence of another amino acid, highlighted the following amino acids as possible substrates for TDH: L-histamine, L-isoleucine, L-lysine and L-leucine. All of these amino acids show average activity rates that were less than 2% of the activity exhibited by TDH with 30mM L-threonine. These data suggest that TDH is highly selective for L-threonine, although it may have weak affinity for other amino acids. Another observation of interest is the apparent stimulation of TDH activity by L-glutamine. As this amino acid does not appear to be a substrate for TDH, it raises the possibility that it could stimulate the enzyme by allostery.

3.2.2 Alteration and Optimization of TDH activity

The velocity of TDH activity is higher at 37°C than at room temperature, 23°C. The plots of [substrate] versus rate at each temperature fit to curves with similar profiles and appear to be parallel (see Figure 3.2.5 and Figure 3.2.6). Thus, the increase in rate may be explained simply

by an increase in the kinetic energy of substrates for the reaction as the temperature increases. To confirm this, a rearrangement of the Arrhenius equation:

$$E_a = \ln k_{cat} + \ln\left(\frac{k_B T}{h}\right) \times (RT) \qquad \textbf{26}$$

where E_a is the activation energy for catalysis, k_B is Boltzmann's constant, h is Planck's constant, T is temperature (in Kelvin) and R is the ideal gas constant, was used to calculate very similar E_a values from k_{cat} of 33.015kJ (7.891kcal) at 37°C (310K) and 31.375kJ (7.523kcal) at 23°C 296K. The difference in activation energy is not significant. These results suggested that inhibitory assays and screening could be carried out at either temperature.

The methods for various investigations of factors affecting TDH activity are described in Section 2.2.2. The results of those investigations are described below. There was a linear relationship between the concentration of the TDH enzyme and the rate of catalysis at saturating substrate concentrations. This implies that the catalytic eficiency is not dependent on the enzyme concentration, at least across the range of concentrations that provide an appropriate signal for measurement.

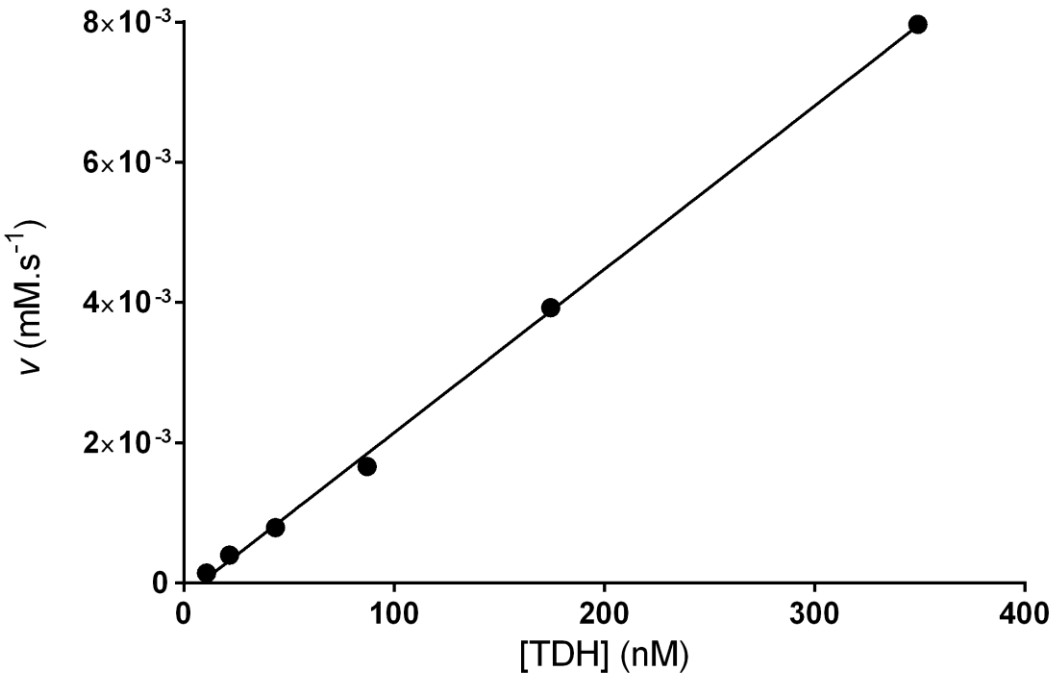


Figure 3.2.9 – a fit of *v* against TDH concentration. All points lie very close to a line fit by linear regression (R^2 = 0.9989).

TDH was found to catalyse at higher rates in the alkaline pH range (see Figure 3.2.10). At pH 6, TDH activity is very low, relative to higher pH levels. The rate of activity reaches its peak at an optimum that lies approximately between pH 8.5 and 9.5. Above pH 9.5, the rate of activity lowers significantly, but appears to plateau between pH 10.5 and 11.5. For this reason, the pH of the reaction buffer for enzyme assays was maintained at pH 8.5. Measurement of the rate of TDH activity was not possible at pH 12.5 and above, as a spontaneous increase in absorbance at 340nm, presumably caused by breakdown of NAD^+, was observed under these conditions.

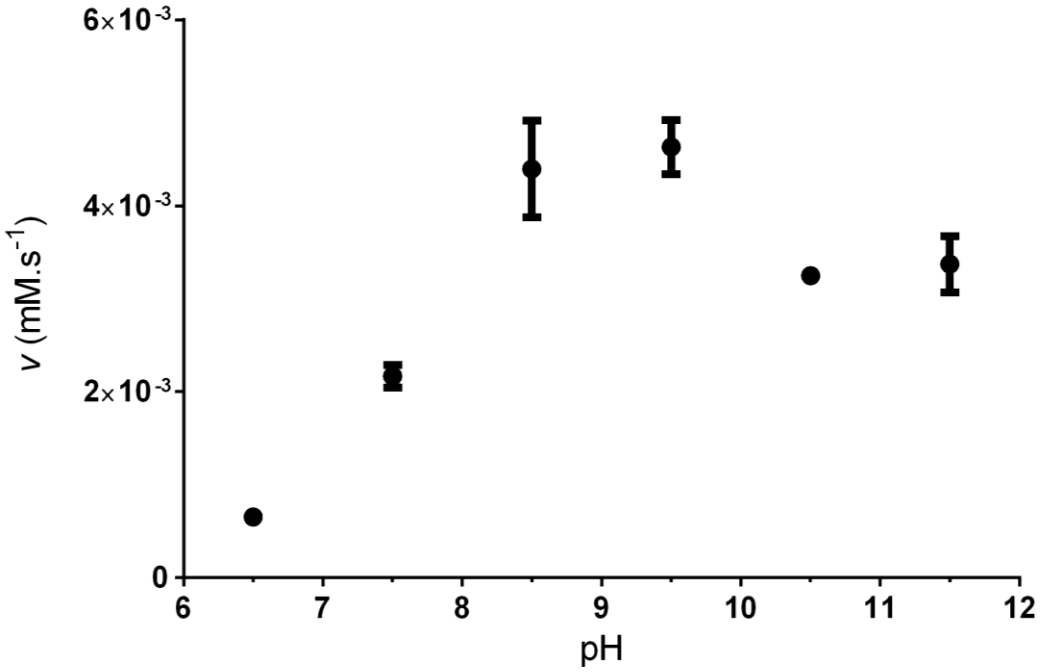


Figure 3.2.10 - a plot of *v* against pH. The optimum pH for TDH activity appears to lie in the region of pH 8.5 - 9.5.

The presence of other additives in the reaction buffer was shown to have no significant effect on TDH activity. In assays where TDH was added to the rest of the reactants to initiate the reaction, the presence of glycerol, up to concentrations of 200mM, and Triton X-100, at concentrations up to 0.1% v/v had no significant effect on TDH activity.

As most enzymatic assays were carried out using an enzyme carrying a poly-histidine tag (His-tag), some TDH with the His-tag removed by thrombin cleavage (see Methods) was tested to ensure that the His-tag did not have any appreciable effect on the kinetic characteristics of TDH.

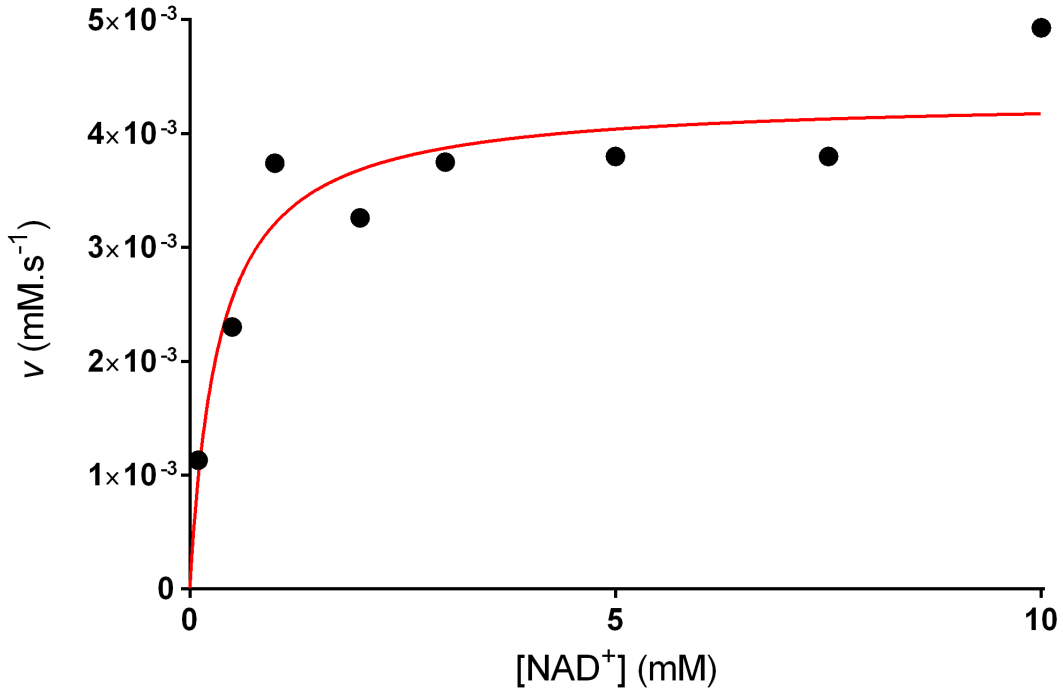


Figure 3.2.11 - the velocity of catabolism by de-tagged TDH, as a function of NAD^+ concentration. A curve describing the Michaelis-Menten equation is fitted to the data.

The relationship between L-threonine concentration and reaction velocity also appeared to be similar to that shown in Figure 3.2.8, which indicated that the presence of the His-tag was not an underlying cause behind this atypical kinetic behaviour. Figure 3.2.13 shows how the data may be described well by equation 25.

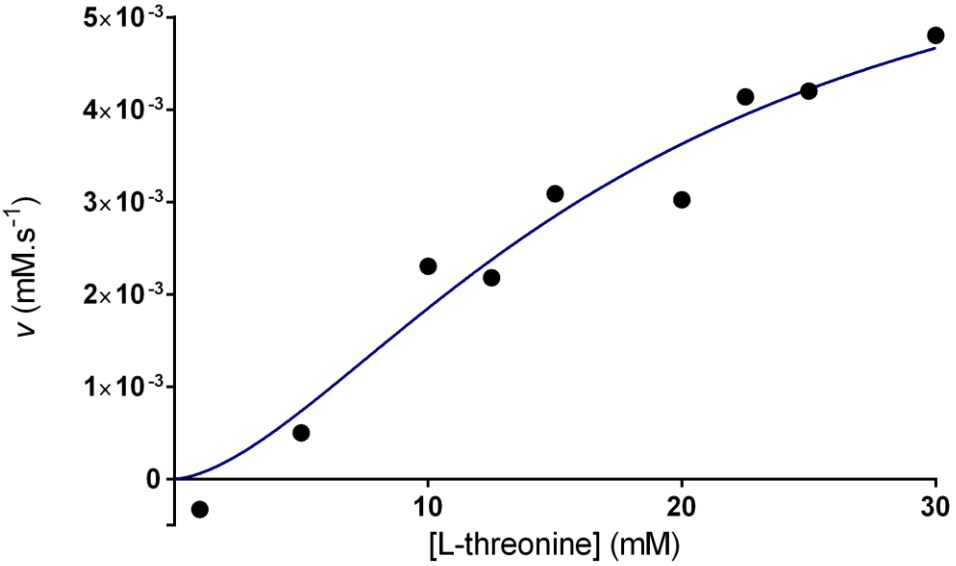


Figure 3.2.12 - plot of de-tagged TDH enzyme velocity against L-threonine concentration. A curve described by the Hill-modified Michaelis-Menten equation (Equation 11) was fitted to the data.

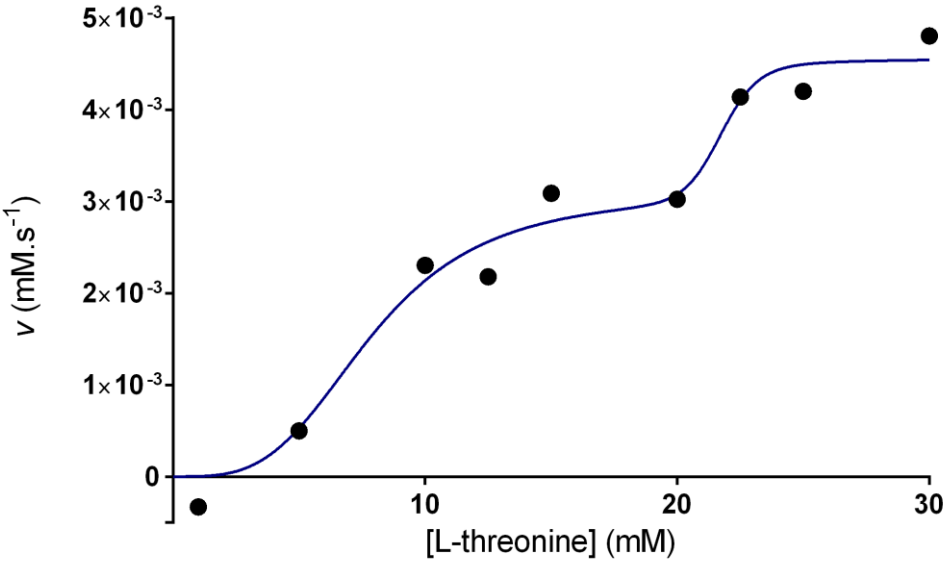


Figure 3.2.13 - plot of de-tagged TDH velocity against L-threonine concentration. A curve described by equation 25 was fitted to the data.

By measuring the effects of different monovalent ions, divalent ions and covalent modifiers on an enzyme's activity, one can gain useful information on important structural features and functional groups. Table 3.2.2 and Table 3.2.3 list the results of assays that measure the effect on TDH activity of incubating the enzyme with various metal salts, covalent modifiers and chelators.

Table 3.2.2 - the effect of different metal salts and covalent modifiers on TDH activity. *The inhibitory activities of N-ethylmaleimide and iodoacetamide were determined by measuring activity at increasing incubation times (see Figure 3.2.14).

Compound	Percentage activity ± SD
Na_2SO_4	108.5% ± 17.7%
$Zn(CH_3CO_2)_2$	-2.3% ± 2.9%
$ZnSO_4$	0.4% ± 5.3%
$CoCl_2$	72.4% ± 16.2%
$MnCl_2$	105.1% ± 8.8%
$CdCl_2$	-1.4% ± 4.0%
$CuCl_2$	-2.0% ± 5.5%
$FeSO_4$	59.3% ± 14.4%
$FeCl_3$	69.3% ± 16.6%
$CaCl_2$	80.8% ± 7.8%
EDTA (5mM)	111.4% ± 11.4%
$AgNO_3$	2.5% ± 3.9%
$HgCl_2$	-1.6% ± 4.8%
N-ethylmaleimide (0.5mM)*	0.0%
iodoacetamide (5mM)*	0.0%

The divalent cations, zinc, cadmium, copper, aluminium and mercury, all eliminated TDH activity at 1mM concentrations. Cobalt and calcium ions also caused inhibition, but were less potent than the other ions. Manganese had no effect on TDH activity at the concentration tested. Incubation of TDH with 5mM EDTA overnight did not inhibit catalytic activity. This suggests that divalent cations are not necessary for TDH activity or stability. Silver nitrate ($AgNO_3$) and mercury (II) chloride were found to cause instant inhibition of TDH. On the other hand, the covalent modifiers N-ethylmaleimide and iodoacetamide caused a gradual inhibition that increased with the time of incubation with TDH (see Figure 3.2.14).

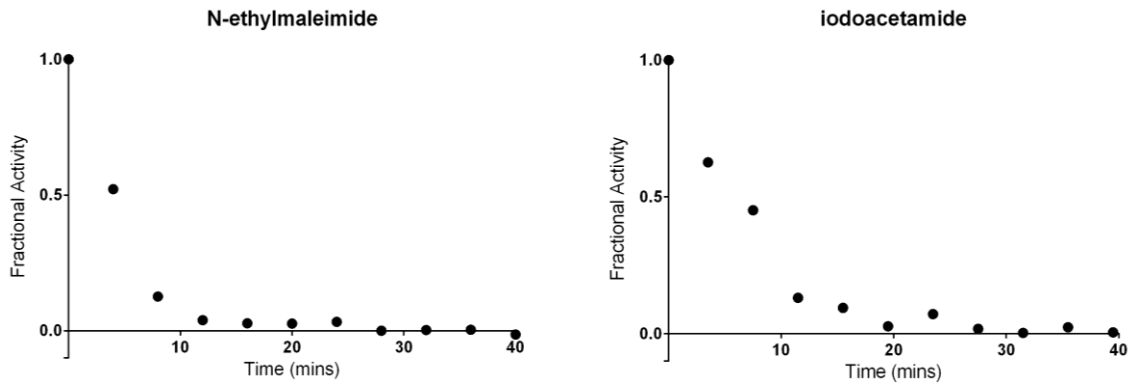


Figure 3.2.14 - the time-dependent inactivation of TDH by N-ethylmaleimide and iodoacetamide.

Table 3.2.3 - the effects of different concentrations of monovalent cations on TDH activity.

Ion (salt)	Concentration	Percentage Activity ± SD
K^+ (KCl)	0.25 mM	107.7% ± 2.8%
	0.5 mM	91.9% ± 2.7%
	1.0 mM	103.6% ± 3.3%
	2.0 mM	122.3% ± 4.9%
	5.0 mM	122.8% ± 16.5%
	10.0 mM	124.2% ± 0.1%
NH_4^+ (NH_4Cl)	0.25 mM	102.1% ± 0.8%
	0.5 mM	105.6% ± 3.5%
	1.0 mM	107.6% ± 1.0%
	2.0 mM	108.4% ± 5.0%
	5.0 mM	123.4% ± 0.5%
	10.0 mM	121.3% ± 3.7%
Rb^+ (RbCl)	0.25 mM	119.0% ± 18.6%
	0.5 mM	133.2% ± 13.5%
	1.0 mM	131.3% ± 16.9%
	2.0 mM	132.8% ± 12.5%
	5.0 mM	151.9% ± 20.9%
	10.0 mM	143.2% ± 27.2%
Cs^+ (CsCl)	0.25 mM	110.5% ± 12.5%
	0.5 mM	112.6% ± 20.3%
	1.0 mM	111.5% ± 12.0%
	2.0 mM	102.1% ± 0.5%
	5.0 mM	120.7% ± 8.2%
	10.0 mM	107.9% ± 7.7%

The data above suggest that potassium (K^+), ammonium (NH_4^+) and rubidium (Rb^+) all increased TDH activity, whilst caesium (Cs^+) had no effect. The stimulation by potassium and ammonium ions is comparable, whilst that of rubidium is slightly higher. These findings indicate that monovalent cations may have a role in increasing the stability of TDH, or that they may play a role in the binding of substrates. In these assays, the monovalent cations may have been competing with or exhibiting an effect that added to that of sodium ions in the reaction buffer.

In assays where TDH was added to the reaction mixture to initiate the reaction, DMSO, at concentrations of up to 10% v/v, was not found to affect activity. When TDH was incubated in reaction buffer with 1-5% v/v DMSO for long periods of time, activity was gradually lost (see discussion of the plate uniformity test and hit validation, below). Therefore, the addition of different components of the storage buffer to the reaction buffer was tested to measure its effects on the enzyme's stability.

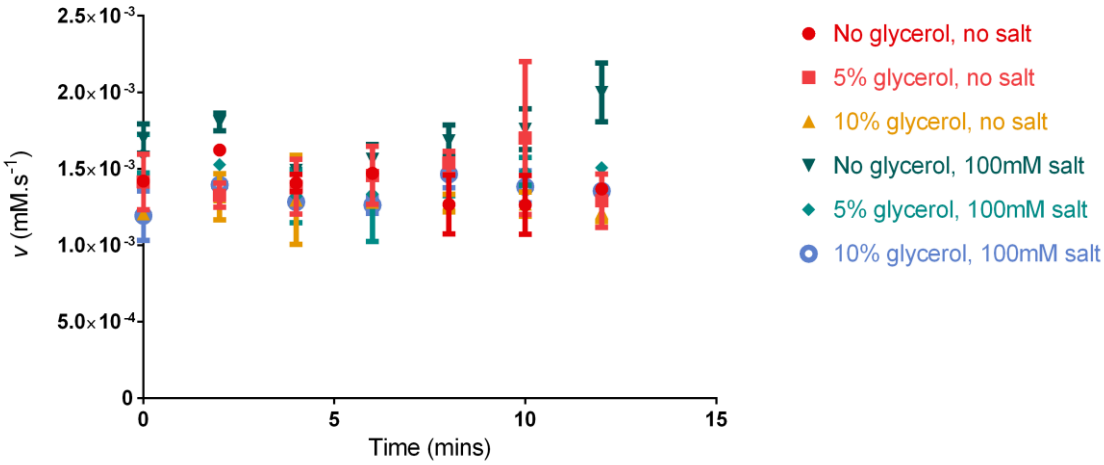


Figure 3.2.15 - TDH activity at room temperature in the presence of different reaction mixture additives. There is no significant decrease in activity as the incubation time in the reaction mixture increases.

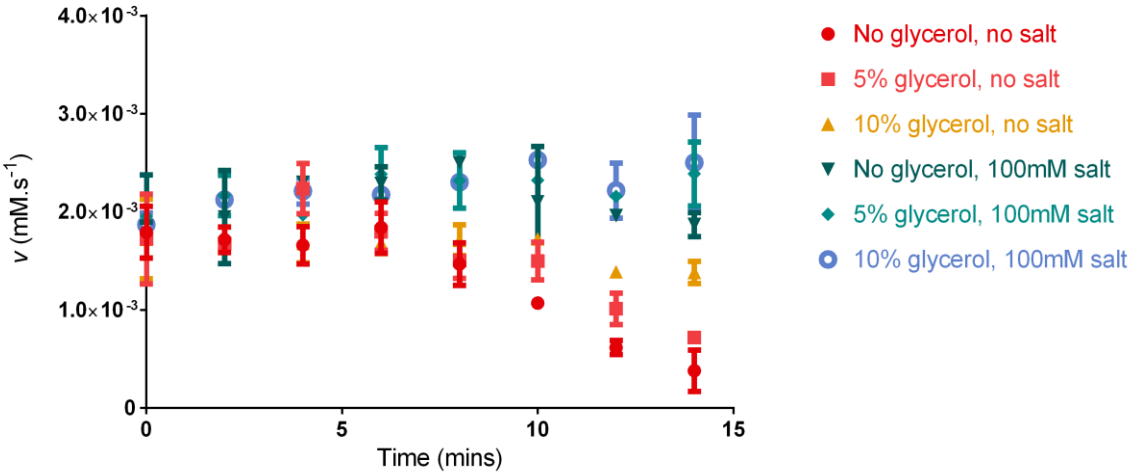


Figure 3.2.16 - TDH activity at 37°C in the presence of different reaction mixture additives. In the absence of salt (100mM NaCl), there is a decrease in the rate of reaction as incubation time increases.

At room temperature, changes to the buffer composition had no effect on TDH activity over the incubation times tested. However, at 37°C, the TDH activity can be seen to decline after incubation times of 6 minutes. Although 10% glycerol may offer some protection of TDH, it is clear that 100mM salt (sodium chloride) was able to preserve TDH activity independently of glycerol concentration. Following on from these results and prior findings on the stimulatory effect of monovalent cations, similar assays were conducted, where TDH was incubated with glycerol and sodium chloride, potassium chloride or rubidium chloride at 37°C. The results of these assays are plotted in Figure 3.2.17.

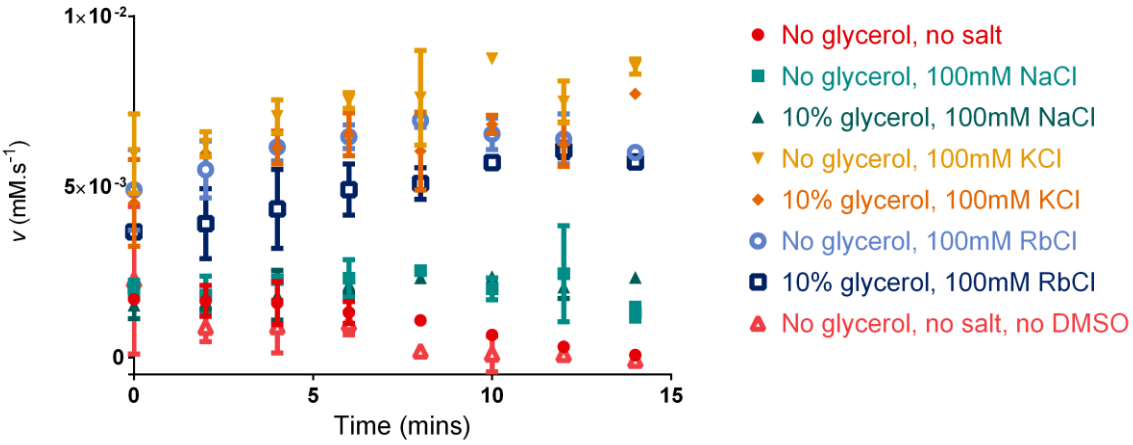


Figure 3.2.17 - TDH activity as a function of incubation time, in the presence of glycerol and different alkali metal salts. NaCl is able to maintain activity over the incubation times measured, whilst added KCl and RbCl greatly increase enzyme activity.

Not only did potassium and rubidium preserve TDH activity over the incubation times studied, they also stimulated activity to more than double that seen in the presence of sodium ions. Control conditions with no glycerol, salt or DMSO showed a similar decline in activity as when there was DMSO, but no glycerol or salt. This suggests that the decline in TDH activity over time is not induced by DMSO.

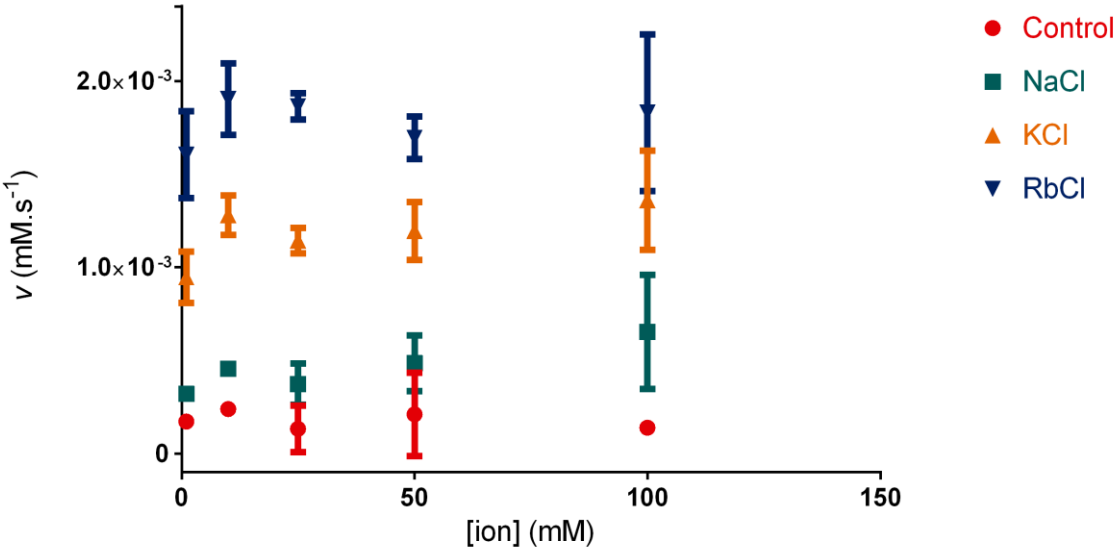


Figure 3.2.18 - the effect of different concentrations of NaCl, KCl and RbCl on TDH activity.

Figure 3.2.18 shows the effect of different cation concentrations on TDH activity. As in the previous assays, 100mM potassium and 100mM rubidium increased activity to similar extents. KCl is more likely to play a physiological role in TDH activity, and the increase in activity from 1mM ion concentration to 100mM was judged to be small, so it was decided that 1mM KCl would be added to the reaction buffer for *in vitro* screening assays. This also allowed a smaller

change to be made to the composition of the reaction buffer, thus reducing the chance of any unforeseen effects on other reaction mixture constituents.

The NAD^+ concentration required to reach half-maximal and maximal TDH rates was considered to be high for an enzyme co-factor. The relationship between L-threonine concentration and rate was also unusual, so measurements of rate were made across a range of substrate concentrations in the presence of 100mM NaCl and 100mM KCl in the TDH storage buffers, and 1mM NaCl and 1mM KCl in the reaction buffers. The profiles of the resulting plots were unchanged from previous assays and were similar with either NaCl or KCl present.

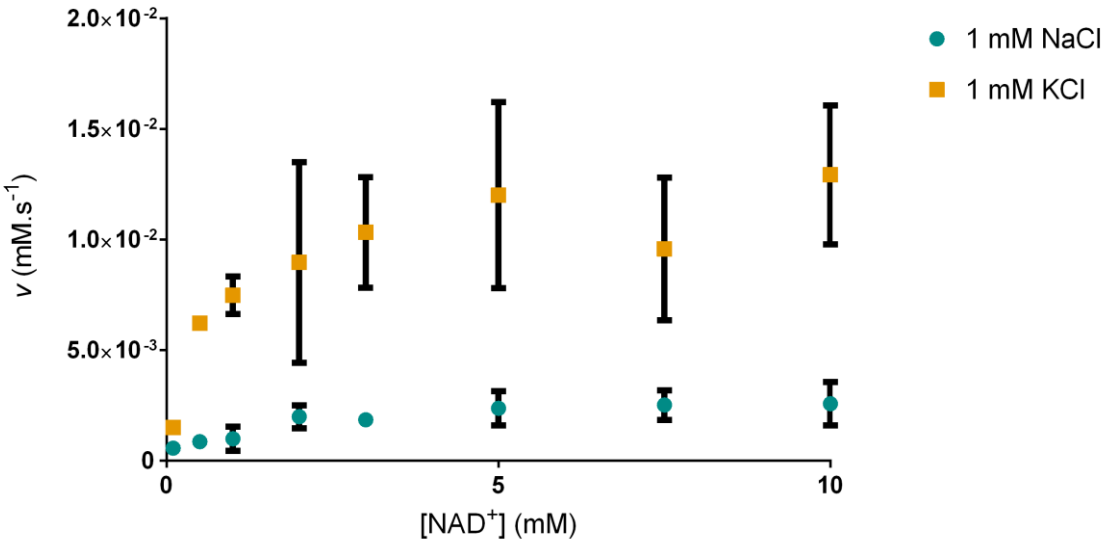


Figure 3.2.19 - the velocity of TDH catalysis as a function of NAD^+ concentration, in the presence of NaCl or KCl.

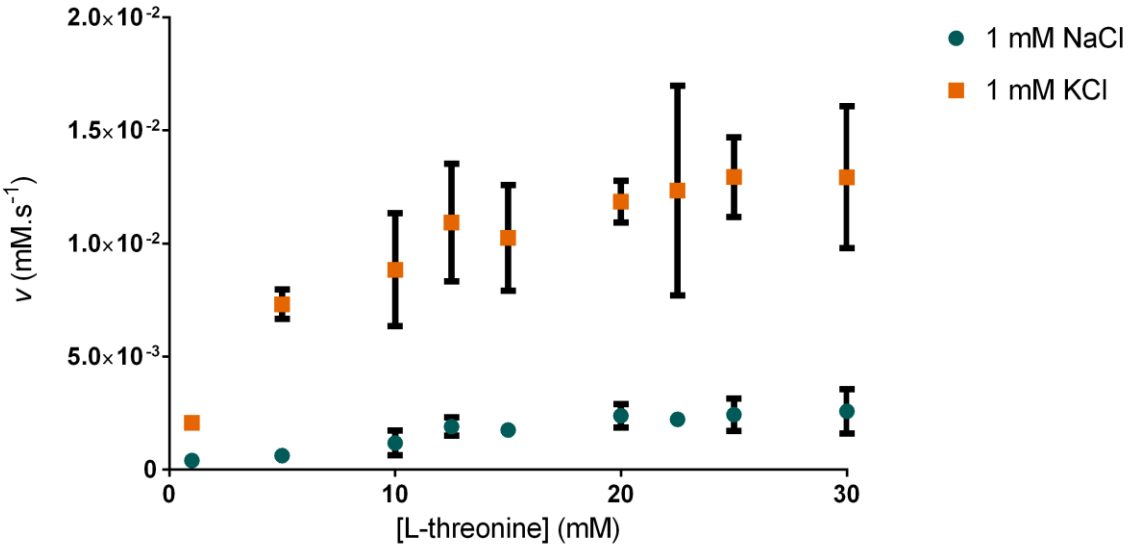


Figure 3.2.20 - the velocity of TDH catalysis as a function of L-threonine concentration, in the presence of NaCl or KCl.

Lineweaver-Burk plots of the same data produce very different slopes, with the points representing data collected from KCl- containing conditions producing a line of best fit with a smaller slope. On a Lineweaver-Burk plot, the equation representing a line of best fit

$$y = mx + c \quad \textbf{27}$$

where *m* is the gradient or slope of the line and c is the *y*-intercept, is represented by a rearrangement of the Michaelis-Menten equation:

$$\frac{1}{v} = \left(\frac{K_M}{V_{max}}\right)\left(\frac{1}{K_M}\right) + \left(\frac{1}{V_{max}}\right) \quad \textbf{28}$$

Therefore, the effect of different assay conditions on enzyme activity can be measured by observing the slope, which is equal to K_M/V_{max}, the y-intercept, which is equal to $1/V_{max}$, and the *x*-intercept, which is equal to $1/K_M$.

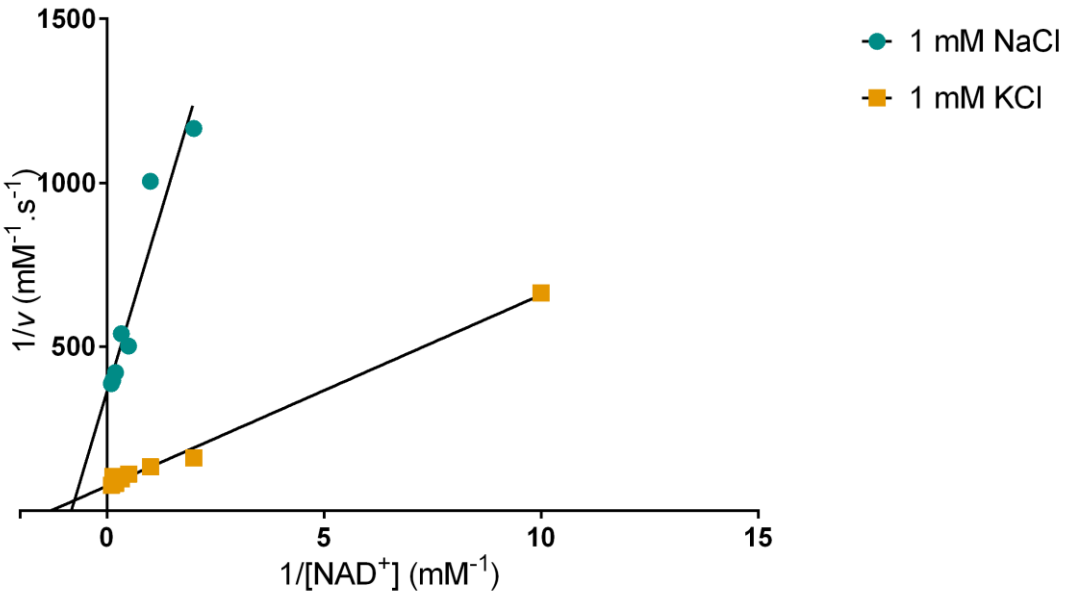


Figure 3.2.21 - Lineweaver-Burk plot of data in Figure 3.2.19.

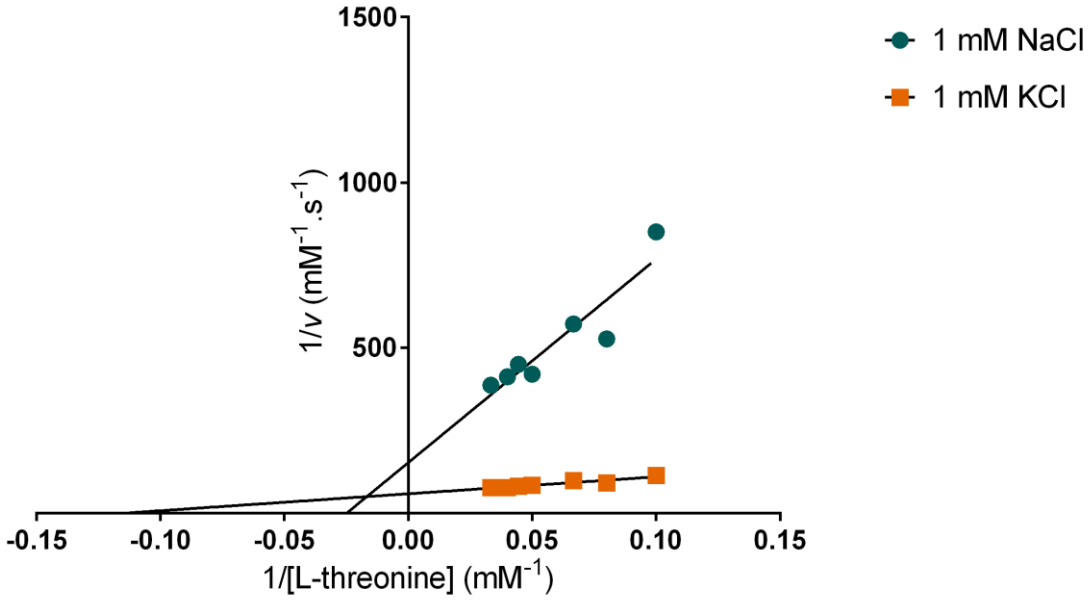


Figure 3.2.22 - Lineweaver-Burk plot of data in Figure 3.2.20. Data points corresponding to lower substrate concentrations are omitted.

The lines in Figure 3.2.21 intersect close to the *x*-axis, but their *y*-intercepts differ. This pattern suggests that the potassium has no effect on NAD^+ binding, as the V_{max} is increased, but the K_M (as predicted by the *x*-intercepts) is roughly the same. Conversely, the *y*-intercepts of each line in Figure 3.2.22 are different, but the apparent K_M values, as would be predicted by observing the *x*-intercept are also different, and this suggests that K^+ has an effect on L-threonine binding. It appears that potassium may aid L-threonine binding, as it decreases the concentration required to reach maximal inhibition. More data is required to determine if this pattern holds over a range of KCl concentrations.

3.2.3 TDH inhibition

3.2.3.1 Rapid-reversible inhibition

The modes of inhibition (MOI) of different TDH inhibitors were estimated by measuring rates of reaction over ranges of substrate concentrations, in the presence of fixed concentrations of inhibitor (see Section 2.2.3 for a detailed description of methods). The data were fitted to different inhibitory equations using the data analysis software Prism. The L-threonine analogues, L-*allo*-threonine, pyruvate and methylglyoxal, and their MOIs are discussed below.

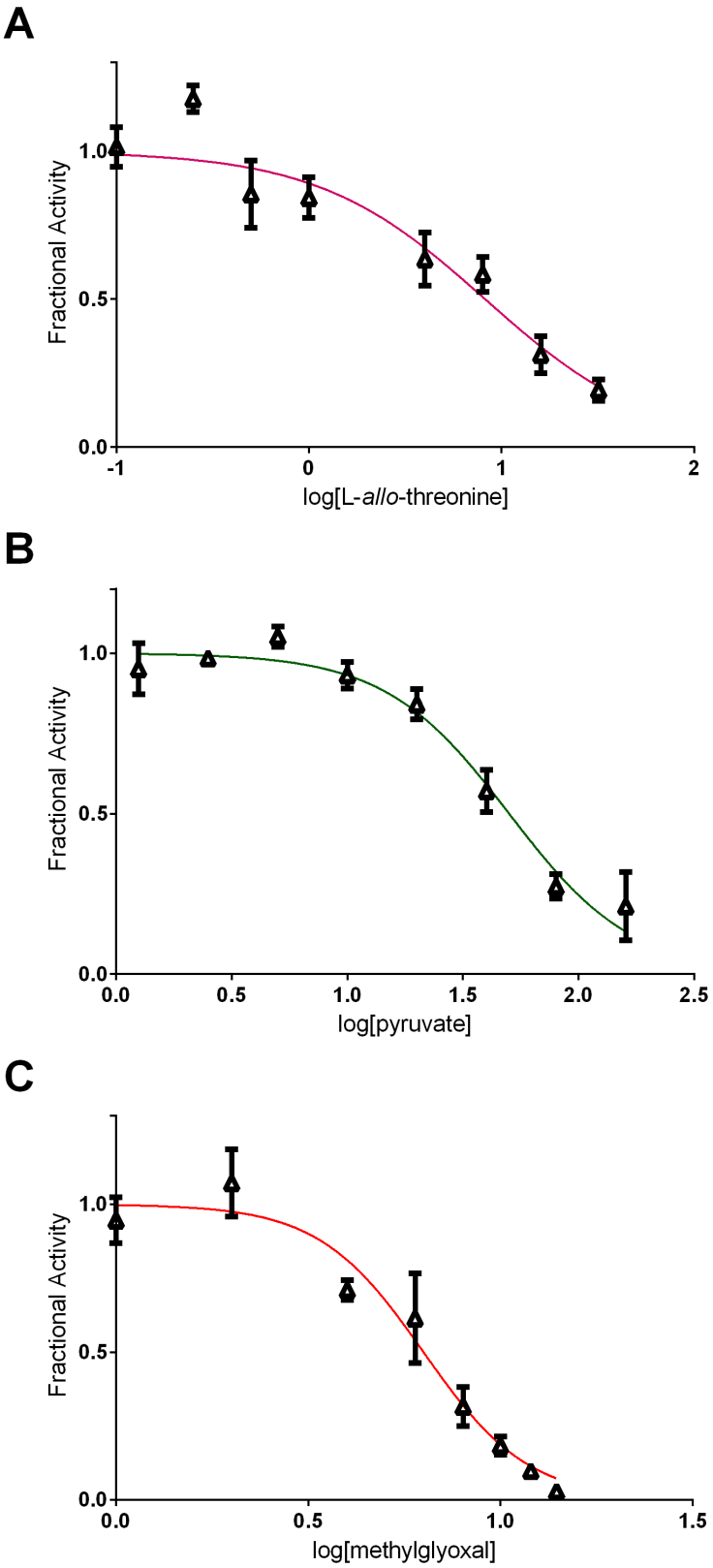


Figure 3.2.23 - dose-response curves fit to data representing fractional enzyme activity as a function of log[inhibitor] concentration. A - the effect of [L-*allo*-threonine] on TDH activity; B - the effect of [pyruvate] on TDH activity; C - the effect of [methylglyoxal] on TDH activity.

L-*allo*-threonine inhibited TDH with an IC_{50} of 8.2mM, whilst pyruvate demonstrated weaker activity and inhibited TDH with an IC_{50} of 50.1mM. Methylglyoxal demonstrated the most potent inhibition of TDH, exhibiting an IC_{50} of 6.3mM. The curve in Figure 3.2.23A does not suggest that L-*allo*-threonine binding is cooperative and a Hill coefficient of 1.00 was calculated for the slope. For pyruvate, a curve with a Hill coefficient of 1.6 was fit, suggesting that there may be some cooperativity involved in pyruvate binding, although this is not clear. In contrast, the curve calculated for the plot of methylglyoxal concentration against rate had a Hill coefficient of 3.2, clearly showing that its action was cooperative. Further investigations of the modes of inhibition (MOIs) of these compounds are discussed below.

L-*allo*-threonine

The data indicates that L-*allo*-threonine inhibition is uncompetitive in relation to NAD^+. This means that it preferentially binds the TDH-NAD^+ complex. When models representing competitive, uncompetitive and mixed/noncompetitive inhibition were fit to the data, the model with the best fit was for uncompetitive inhibition, and the global R^2 value for all curves was 0.946. When fit to the model for mixed/noncompetitive inhibition (see equation 15), the data analysis software calculated an extremely small alpha value of ~$5x10^{-4}$ and a very high K_i of 14.71M. This points very strongly to an uncompetitive MOI, and this model also fit the data with an R^2 of 0.946. An αK_i of 7.8mM was calculated using the uncompetitive model.

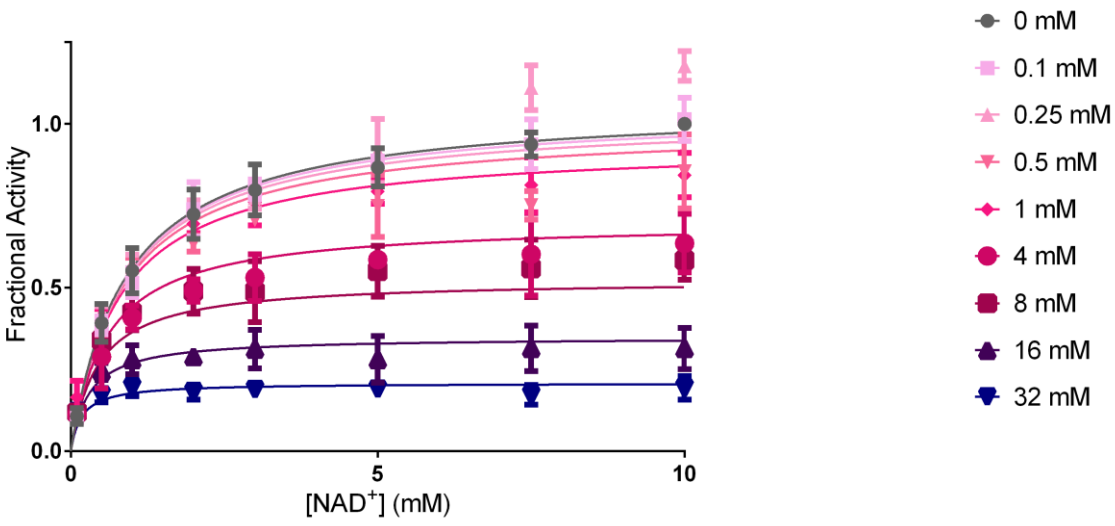


Figure 3.2.24 - normalised TDH activity (fractional activity) as a function of NAD^+ concentration, in the presence of different concentrations of L-*allo*-threonine. Curves described by an uncompetitive MOI (see equation 14) were fit to the data.

As the models used to describe MOI are based on Michaelis-Menten kinetics, they were not ideal for analysing the data on L-threonine concentration and the rate of activity. The data analysis software Prism was able to fit curves based on these models to high R^2 values (>0.9), but they tended to overestimate the value of V_{max}. Despite this, as any particular MOI is defined by its effects on K_M and V_{max}, the data were fit to each equation and compared to determine the

most likely MOI, as was done for the data describing NAD^+ concentration and rate. The competitive inhibition model was the model that best described the inhibition in relation to L-threonine. The mixed model of inhibition was used to calculate a high alpha value of 62, which is in support of a competitive MOI ($\alpha \geq 1.0$ indicates competitive inhibition). A K_i of 1.6mM was calculated for L-*allo*-threonine.

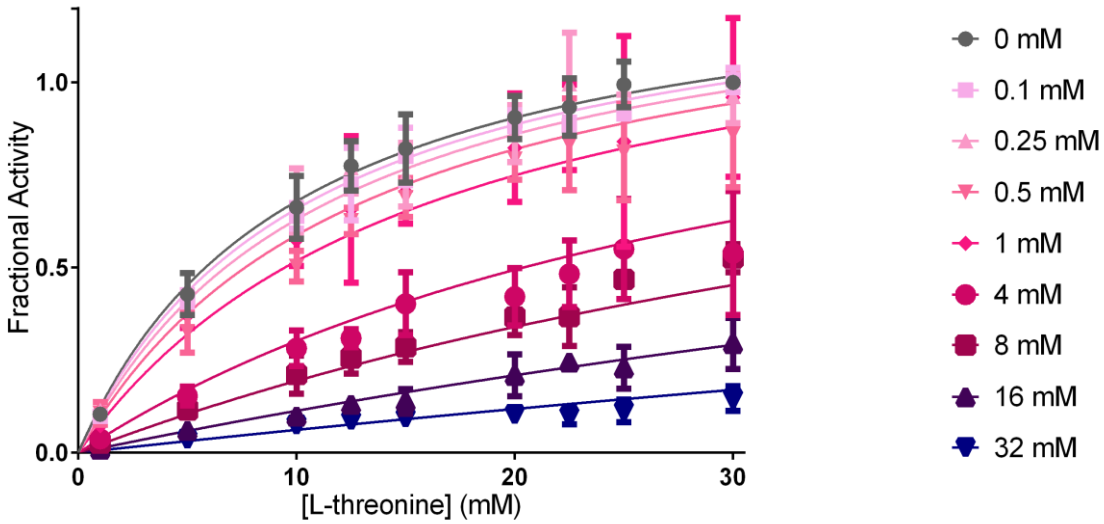


Figure 3.2.25 - plot of fractional activity against L-threonine concentration, measured at different fixed concentrations of L-*allo*-threonine. Curves describing a competitive MOI (see equation 13) are fit to the data.

The MOIs of L-*allo*-threonine that were calculated from enzyme inhibitory data are entirely consistent with the binding mode revealed by x-ray crystallography. As the structure of the compound is so similar to L-threonine, it is not surprising that it preferentially binds to NAD-bound TDH. For the same reason, one would expect it to compete directly with L-threonine for binding.

Pyruvate

Both the noncompetitive and uncompetitive models of MOI fit closely to the data describing the relationship between NAD^+ concentration, rate and pyruvate concentration. The R^2 values were 0.944 and 0.955, respectively. The mixed model of inhibition was the best fit for the data, with an R^2 of 0.947. This model calculated an alpha of 0.134, indicating an uncompetitive mode of inhibition. However, the upper 95% confidence limit for alpha was 0.385 and some researchers stipulate that the upper confidence limit of alpha should be 0.25 for an uncompetitive inhibitor(274). To determine whether the uncompetitive model or noncompetitive model was more appropriate, the two models were compared by calculating values for Aikake's Information Criterion (AIC) using the data analysis software Prism. AIC values for each of the compared models were compared to determine which model was most likely to generate the data. This

calculation determined that the uncompetitive model was more likely to be correct (99.8% probability). The same analysis was performed to ascertain whether the most appropriate model was a mixed noncompetitive-uncompetitive model or an uncompetitive model and the software indicated that the uncompetitive model was more likely to be correct, with a 65.5% probability of having produced the data over the mixed model.

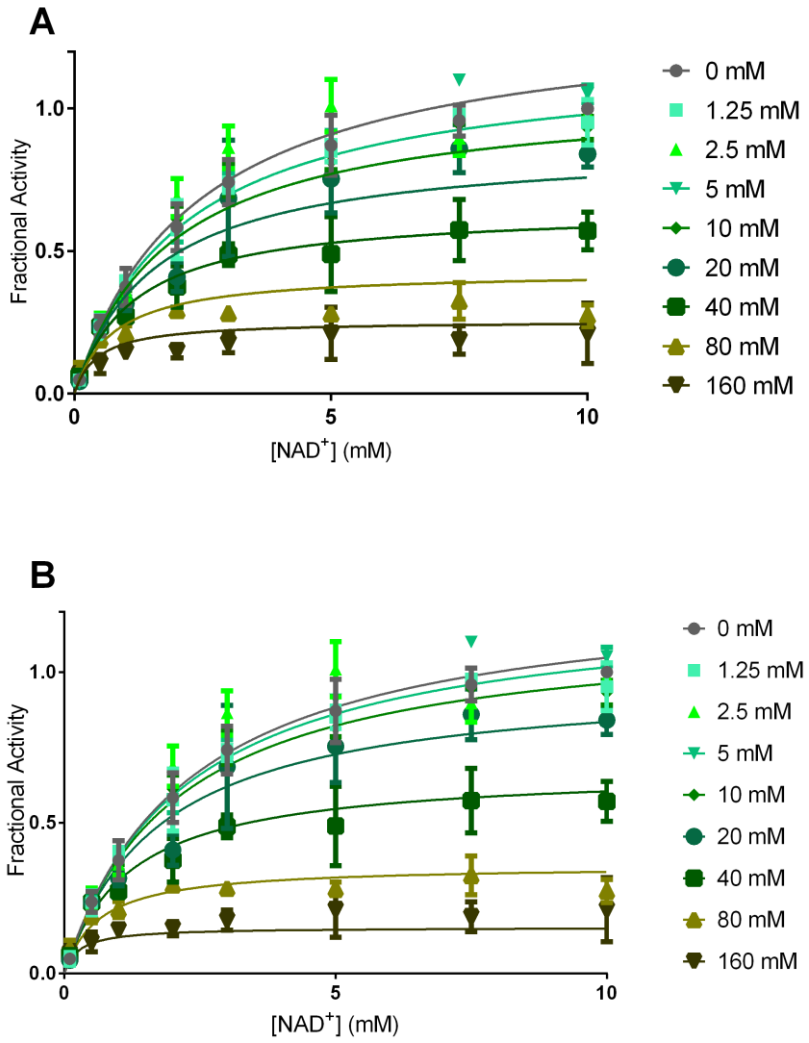


Figure 3.2.26 - plots of fractional activity as a function of NAD^+ concentration, at fixed concentrations of pyruvate. The data are fit to curves describing uncompetitive inhibition (A) and cooperative uncompetitive inhibition (B), described by a modification of equation 14 that adds the Hill coefficient (see equation 30). The fit of the curves to the data collected at 80 and 160mM pyruvate are slightly improved in the graph on the right.

In graph A in Figure 3.2.26, it is evident that the data does not fit well to the data collected at 80mM and 160mM. This can be explained by the fact that the curve relating pyruvate concentration to fractional activity had a Hill coefficient of 1.6, suggesting that there may be some cooperativity involved in pyruvate binding (see Figure 3.2.26B). To take account of these effects, the parameters relating to the inhibitor in the MOI equations were modified with the Hill coefficient, as in the Hill equation (or the Hill modification of the Michaelis-Menten equation).

The Hill modification of the equation for mixed model inhibition (equation 15) can be represented as:

$$v = \frac{V_{max}[S]}{[S]\left(1 + \frac{[I]^h}{\alpha K_i{}^h}\right) + K_M\left(1 + \frac{[I]^h}{K_i{}^h}\right)} \quad \textbf{29}$$

Similarly, the Hill modification of the model for uncompetitive inhibition (see equation 14) can be written as:

$$v = \frac{V_{max}[S]}{[S]\left(1 + \frac{[I]^h}{\alpha K_i{}^h}\right) + K_M} \quad \textbf{30}$$

Similar modifications to inhibitory equations can be found in the literature(275). To ensure that any improved fits of the data were solely due to the addition of an extra parameter, an F test was used to compare the new models to the original models. The F test calculates whether improved fits to data in more complex equations are statistically significant or just the result of having a higher number of degrees of freedom. The test confirmed that the improved fits to the data (R^2 = 0.961 for Hill modifications of uncompetitive and mixed models) were statistically significant. Now, the MOI was more easily assigned as uncompetitive inhibition. The Hill-modified mixed-model calculated a very low alpha value of 0.0014. The Hill-modified uncompetitive model calculated an αK_i of 42.8 and Hill coefficient of 1.52, which is very similar to the Hill coefficient of 1.6 that was calculated for the curve relating pyruvate concentration to rate (see Figure 3.2.23B).

Fits of the data from studies conducted at different L-threonine concentrations to different models of MOI strongly indicated that pyruvate acted competitively. A mixed-inhibition model fit with a R^2 value of 0.945 and a calculated alpha of 66.0, strongly suggesting that the MOI is competitive. A competitive inhibition model was fit to the data, and produced an equivalent R^2 value. This model yielded a K_i of 17.1mM, which is similar to the value predicted by the mixed-model, 17.6mM. A Hill-modified mixed model slightly improved the fit to the data, producing an R^2 of 0.946. A very high alpha, indicative of competitive inhibition, was predicted again, and a modest value for *h* of 1.13 was calculated, suggesting that there was no cooperative behaviour or that cooperativity was only marginal. This model predicted a slightly higher K_i of 19.5mM.

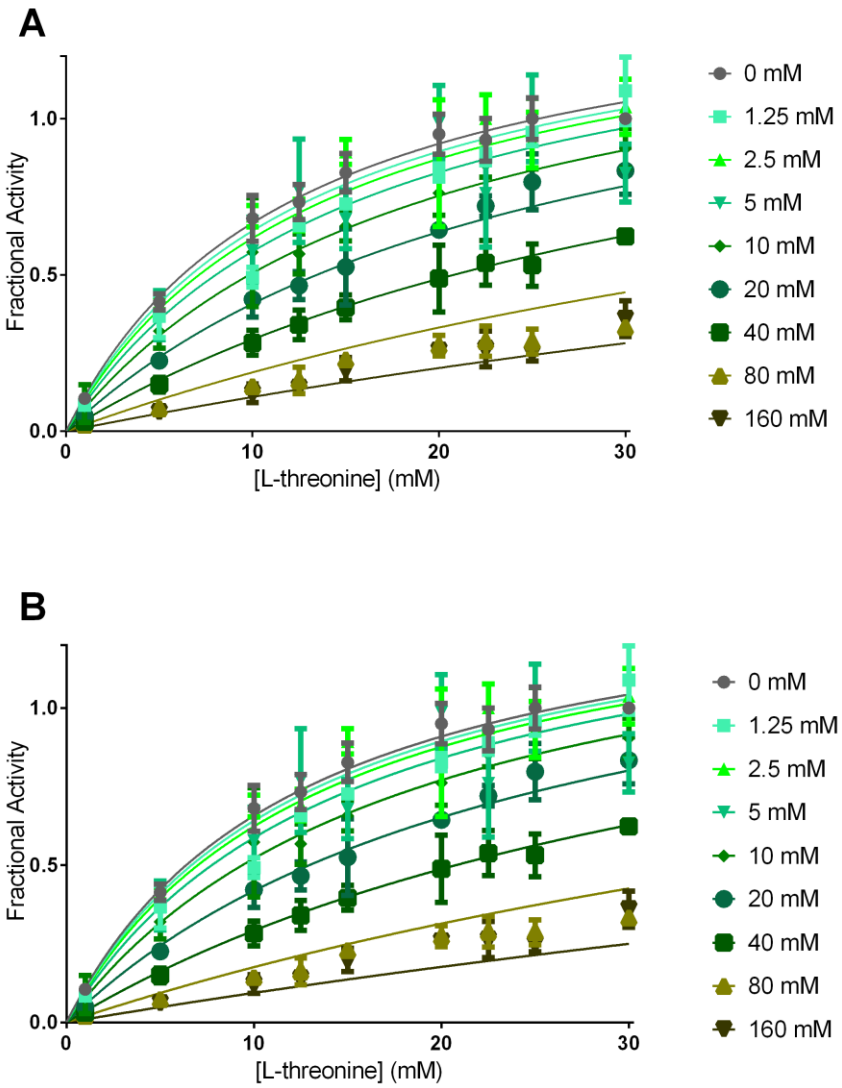


Figure 3.2.27 - plots of fractional activity against L-threonine concentration measured in the presence of different concentrations of pyruvate. Curves were fit to the data using the mixed inhibition model (A) and the cooperative mixed inhibition model (B). The fits to the data are very similar and both models indicated that the MOI of pyruvate was competitive in relation to L-threonine.

Once again, the results of these MOI analyses are consistent with structural data on the binding of pyruvate: pyruvate competes with L-threonine for its binding site. Like L-threonine, pyruvate binding is also dependent on the presence of NAD^+, thus it has a strong preference for binding to the *holo* form of TDH.

Methylglyoxal

As discussed earlier, the cooperativity of methylglyoxal inhibition was much more pronounced than that displayed for pyruvate. Standard MOI equations could not be fit to data for all the methylglyoxal concentrations tested. Therefore, using Hill-modified versions of the standard

equations, it was determined that methylglyoxal acted noncompetitively in relation to NAD^+. The Hill-modified mixed model of inhibition fit the data with an R^2 of 0.976 and predicted an alpha of 1.15 with the 95% confidence interval 0.5517 to 1.748. This was highly suggestive of a noncompetitive model. This was confirmed by fitting the data to a Hill-modified version of equation 15, which assumes an alpha value of 1.0 (indicating an equal affinity for free and substrate-bound enzyme):

$$v = \frac{V_{max}[S]}{([S] + K_M)\left(1 + \frac{[I]^h}{K_i{}^h}\right)} \qquad \textbf{31}$$

This equation also fit the data with an R^2 of 0.976. A K_i of 6.1mM was calculated and the Hill coefficient was 3.3, which is almost identical to that observed for the slope of methylglyoxal concentration and fractional activity (see Figure 3.2.23C).

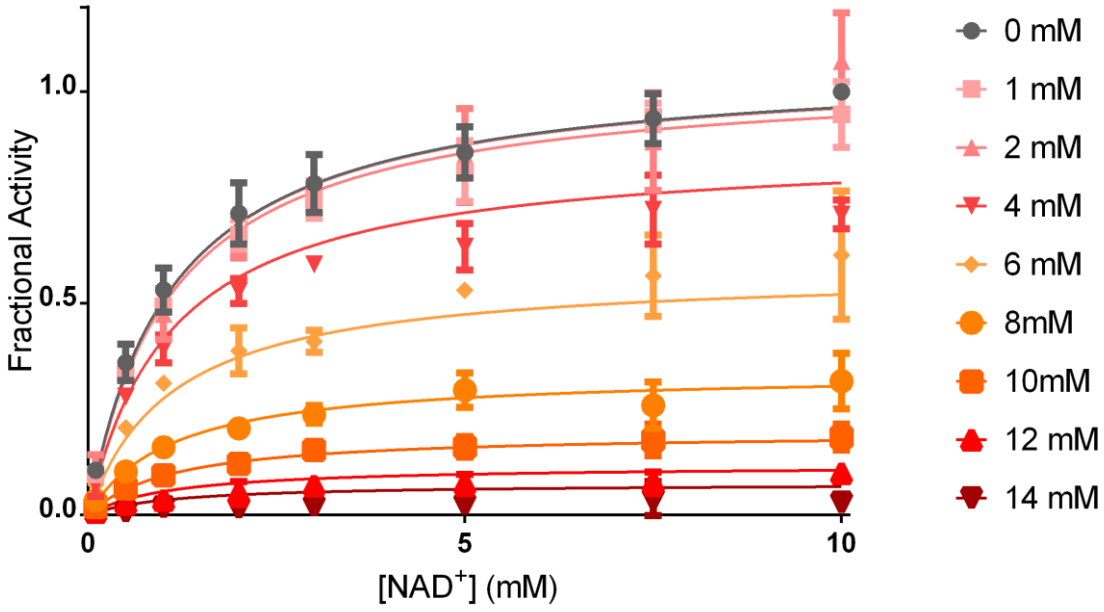


Figure 3.2.28 - a plot of Fractional activity against NAD^+ concentration in the presence of different concentrations of methylglyoxal. Curves described by equation 31, which describes cooperative noncompetitive inhibition were fit to the data.

A fit of equation 29 to the data regarding L-threonine concentration and rate fit with an R^2 of 0.971. An alpha value of 49.0 signalled that the MOI was competitive. A Hill-modified version of equation 13, describing cooperative competitive inhibition:

$$v = \frac{V_{max}[S]}{[S] + K_M\left(1 + \frac{[I]^h}{K_i{}^h}\right)} \qquad \textbf{32}$$

was used to fit the data, producing an R^2 of 0.971. The model was used to calculate a K_i of 4.7mM. Both the mixed and competitive models gave Hill coefficients of 3.4, which are similar

to the Hill coefficient of 3.2 that was calculated for the curve fitted to the plot of methylglyoxal concentration against fractional activity (see Figure 3.2.23C).

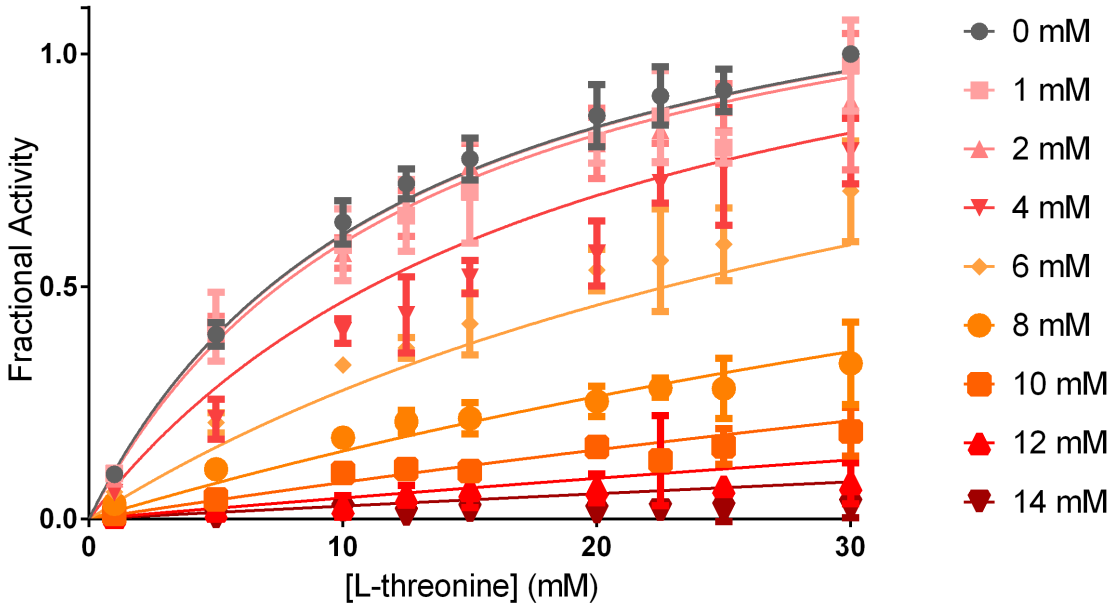


Figure 3.2.29 - plot of fractional activity at different L-threonine concentrations, in the presence of different concentrations of methylglyoxal. Curves were fitted using a cooperative competitive MOI (see equation 32).

The noncompetitive MOI in relation to NAD^+ and the competitive MOI in relation to L-threonine describes a mechanism that involves methylglyoxal interacting with TDH in a way that is mutually exclusive with L-threonine binding, but independent of NAD^+ binding.

The confirmation of MOIs of L-*allo*-threonine, pyruvate and methylglyoxal using enzyme assays demonstrates their utility as a way of predicting the MOI of new inhibitors.

3.2.3.2 Time-dependent inhibition

Inhibition by the human aldehyde dehydrogenase inhibitor, tetraethylthiuram disulphide (TETD or disulfiram), was a gradual inhibition of TDH that increased with the time that it was incubated with the enzyme. Therefore, determining its MOI required different assays and data analysis techniques to the rapid-reversible inhibitors described above. These assays are described in the methods section.

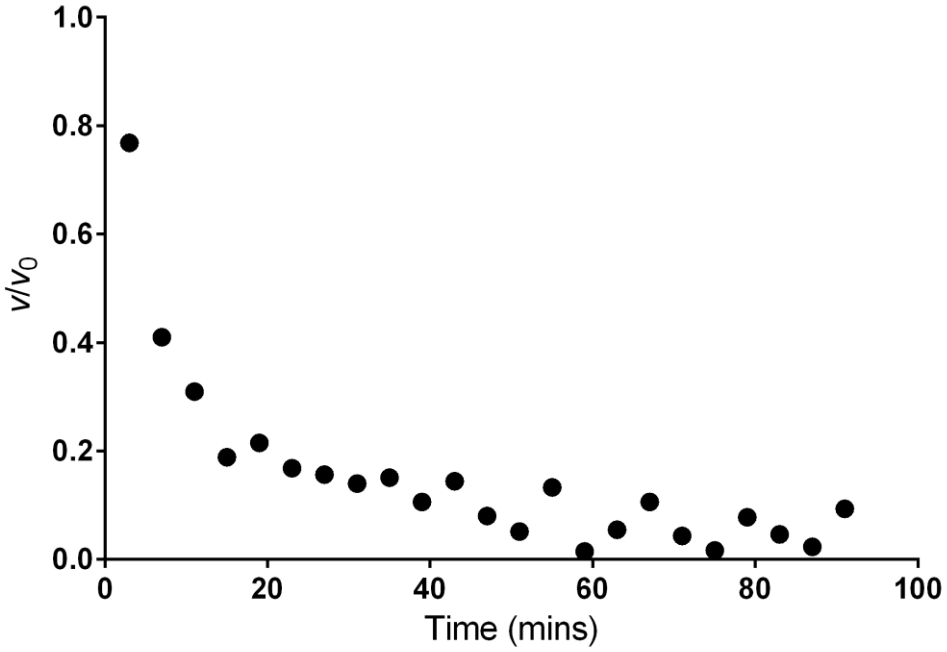


Figure 3.2.30 - plot of fractional activity at increasing incubation times of TDH with TETD.

Data similar to that presented in Figure 3.2.30 were collected over the following TETD concentrations: 90, 175, 250, 350, 525 and 700 µM. The value of k_{obs} was calculated using equation 16, and this value was plotted against the relevant TETD concentration to determine the underlying mechanism behind the time-dependency of inhibition by TETD.

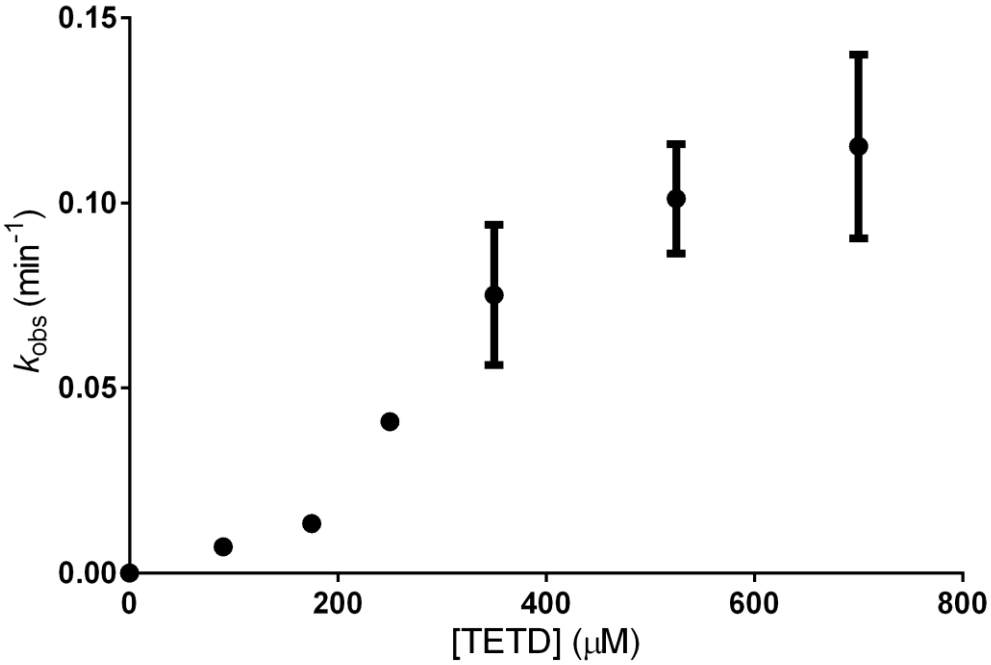


Figure 3.2.31 - plot of k_{obs} as a function of TETD concentration.

The plot in Figure 3.2.31 is sigmoidal, which rules out a reversible slow-binding mechanism, which would manifest in a plot that shows a linear relationship between k_{obs} and inhibitor concentration. Two further mechanisms are outlined by schemes A and B in Figure 3.2.32, below.

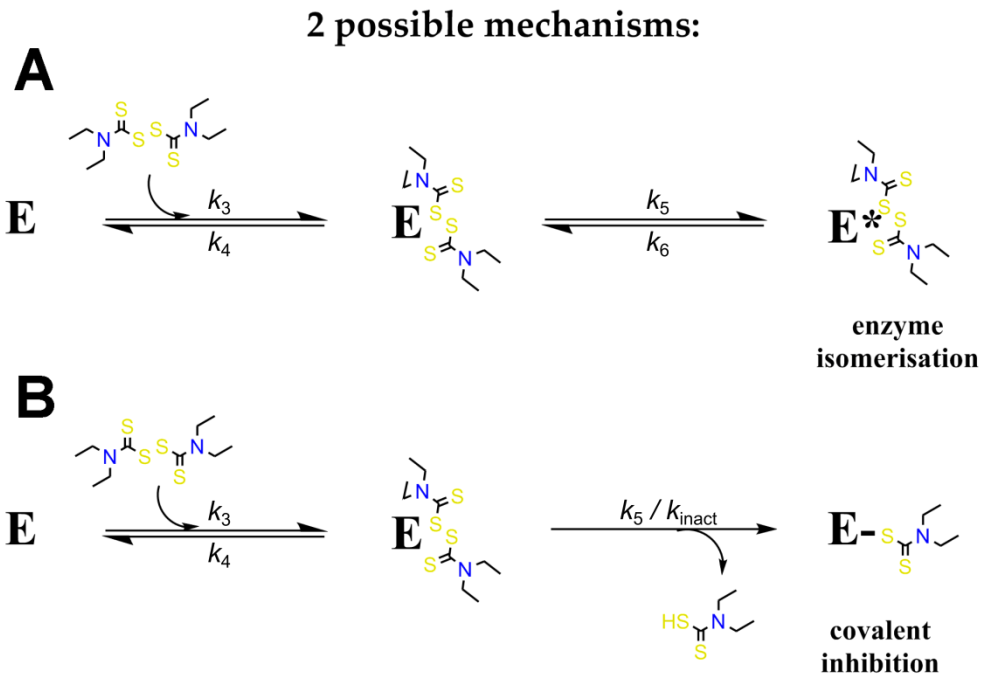


Figure 3.2.32 - possible mechanisms of TDH inhibition by TETD. Scheme A describes an initial binding step, which is characterised by the constants k_3 and k_4, followed by a slower step involving enzyme isomerization, which leads to stronger binding. Scheme B also consists of an initial reversible binding step, but this is followed by irreversible inactivation of the enzyme by covalent modification. This second step is described by the constant for inactivation, k_5 or k_{inact}. Note that if k_6 in scheme A is very small, then the two schemes are indistinguishable from each other.

The plots that would be explained by such binding mechanisms in Figure 3.2.32 should show hyperbolic relationships(194,276). The sigmoidal relationship between TETD concentration and k_{obs} raises the possibility of cooperative binding by TETD. Therefore, the equations describing schemes A and B from Figure 3.2.32 were modified with the Hill coefficient. The resulting equation used for mechanism A was:

$$k_{obs} = k_6 + \left[\frac{k_5[I]^h}{K_i^{app^h} + [I]^h}\right] \quad \textbf{33}$$

The resulting equation describing mechanism B was:

$$k_{obs} = \frac{k_5[I]^h}{K_i^{app^h} + [I]^h} \quad \textbf{34}$$

The relationship between the apparent K_i (K_i^{app}) and the true K_i depends on the MOI of the inhibitor. The data in Figure 3.2.31 were fitted by non-linear regression to equations 30 and 31. The results are shown in Figure 3.2.33, below.

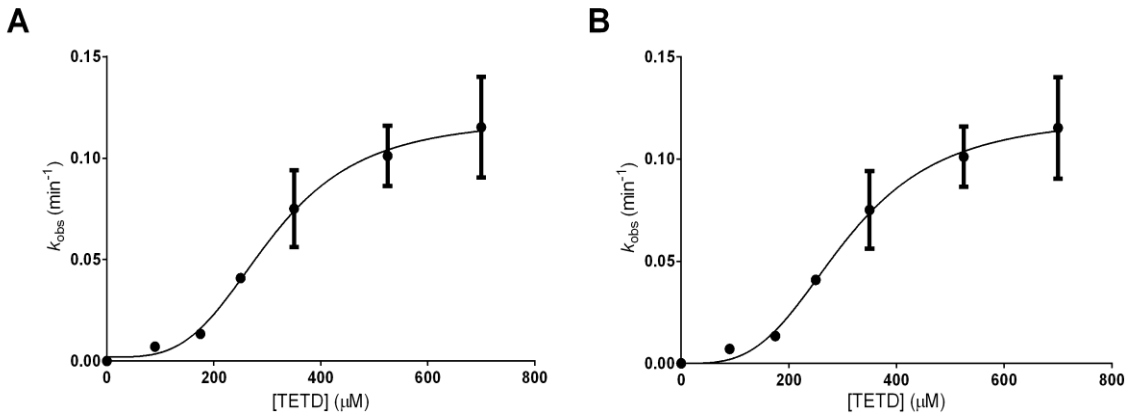


Figure 3.2.33 - plots of k_{obs} against TETD concentration. Curves are fitted to the data using equation 33 (left) and equation 34 (right). The curves fitted equally well to the data, and the *y*-intercept, which corresponds to k_6, is very close to zero in the plot A, which corresponds to scheme A.

The R^2 values for the data fit to equations 30 and 31 are both 0.997. A comparison using AIC highlighted mechanism B as the best fit for the data. Furthermore, the value for k_6 calculated using equation 33 was 0.002, compared to a k_5 value of 0.118. This makes the two mechanisms almost indistinguishable, but it suggests that the final step is virtually irreversible, if not truly irreversible. The apparent K_i (K_i^{app}) is calculated by equation 34 as 307.0μM. The Hill coefficient was equal to 3.3.

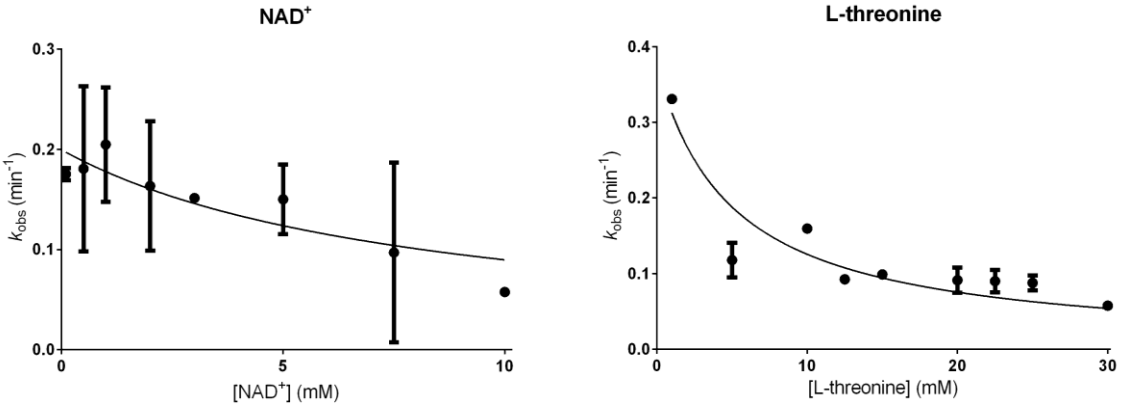


Figure 3.2.34 - plots of k_{obs} as a function of substrate concentration. The data in these plots were obtained by calculating k_{obs} using the time-dependent inhibition observed by incubating 350μM TETD with TDH prior to adding different concentrations of one substrate in the presence of saturating concentrations of the other. The relationships between k_{obs} and substrate concentration are very distinct for different modes of inhibition. An inverse relationship is observed for competitive inhibition, whereas a direct relationship is observed for uncompetitive inhibition. There is no relationship for noncompetitive inhibition. Curves describing competitive time-dependent inhibition (see equation 17) are fit to the data.

To determine the MOI in relation to the substrates L-threonine and NAD^+, a single concentration of TETD was tested in the presence of a range of substrate concentrations. The data show an inverse relationship between k_{obs} and substrate concentration, suggesting an MOI that is competitive towards both NAD^+ and L-threonine binding. Accordingly, the true K_i of TETD may be estimated using the Cheng-Prusoff equation for competitive inhibition, where IC_{50} is replaced with K_i^{app}:

$$K_i = \frac{K_i^{app}}{1 + \frac{[S]}{K_M}} \quad \textbf{35}$$

As the relationship between NAD^+ concentration and rate follows Michaelis-Menten kinetics, using values of 10 mM for [S] (or $[NAD^+]$ and 1.0 mM for K_M, the K_i of TETD is calculated as 27.9 µM. Note that this K_i represents the initial reversible binding of TETD to TDH, as the second step is governed by k_5 (see Figure 3.2.32) and is irreversible. The value for k_5 or k_{inact} for TETD was calculated as $0.12 min^{-1}$.

3.2.4 Preliminary enzymatic studies of TDH from *Clostridium difficile*

The rate of catalysis of TDH from *Clostridium difficile* (*Cd*TDH) was measured across a wide range of NAD^+ and L-threonine concentrations. The relationships between the substrate concentrations and reaction velocity are quite similar to those observed for TDH from *Trypanosoma brucei*.

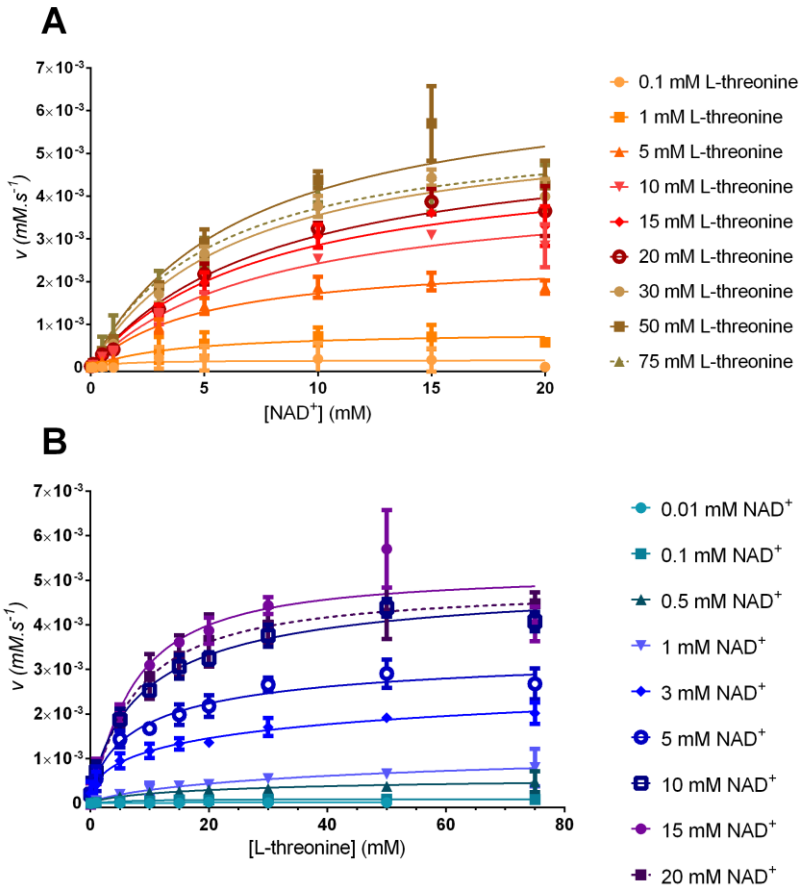


Figure 3.2.35 - plots of *Cd*TDH reaction velocity as a function of substrate concentration, at fixed concentrations of the second substrate (A = NAD^+; B = L-threonine). Curves were fitted using the Michaelis-Menten equation (A) or the Hill modification of the Michaelis-Menten equation (B). Dotted curves are fitted to data where an increase in substrate concentration caused a decrease in reaction velocity, relative to a lower concentration

The Lineweaver-Burk plot of data relating NAD^+ concentration to reaction velocity at different fixed L-threonine concentrations shows a pattern of lines with slopes that decrease as the L-threonine concentration is increased. This is a clear indication of a sequential ordered mechanism of action. The lines intersect near to where they intercept the *x*-axis, suggesting that L-threonine concentration does not affect the binding affinity of NAD^+. In a way that is somewhat similar to the data obtained for TDH from *T. brucei*, the Lineweaver-Burk plot of the data relating rate of activity and L-threonine concentration produces a pattern of lines that show a mixed pattern of decreasing slopes with higher NAD^+ concentrations, and lines that are parallel to each other.

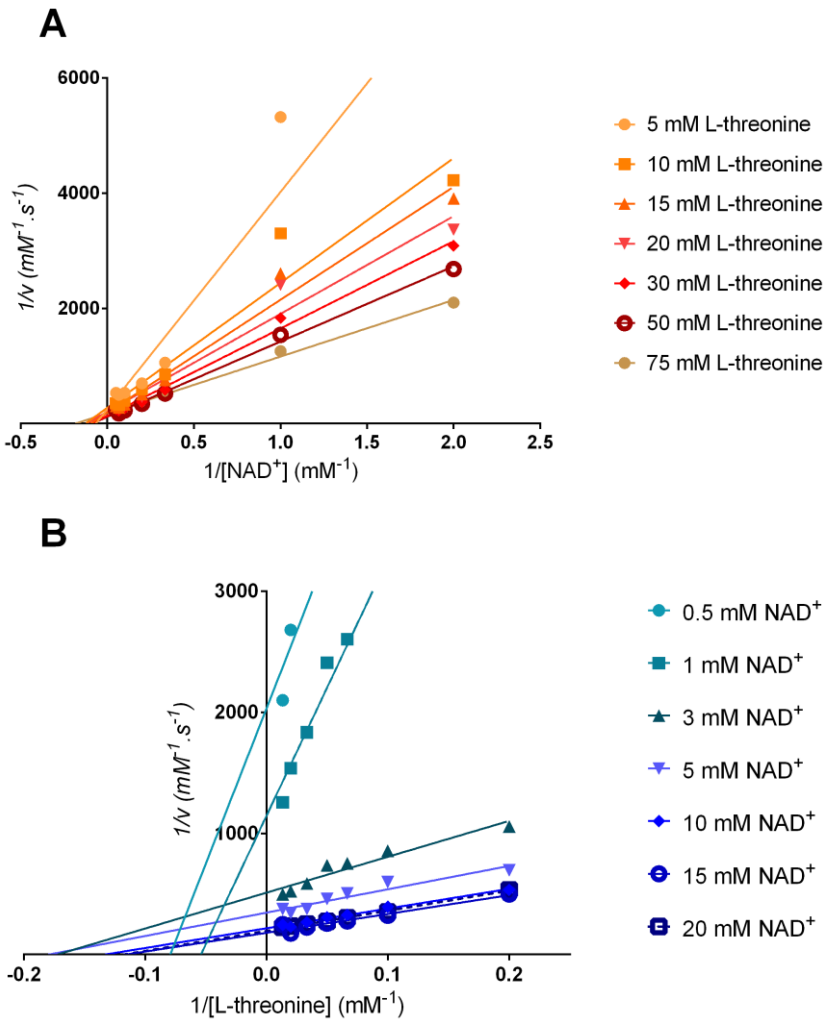


Figure 3.2.36 - Lineweaver-Burk plots of the data plotted in Figure 3.2.35. Dotted lines are fitted to data where an increase in the concentration of the other substrate led to a decrease in reaction velocity.

It is unclear whether the reaction mechanism is compulsory ordered or a random sequential reaction from these data alone. However, the change in *x*-intercept at higher NAD^+ concentration does suggest that NAD^+ may be the first substrate to bind.

The saturating concentrations of NAD^+ and L-threonine appear to be higher for *Cd*TDH, than for TDH from *T. brucei*. The saturating NAD^+ concentration appears to be high, at 15mM. The relationship between the NAD^+ concentration follows Michaelis-Menten kinetics, and a K_M of approximately 7mM is estimated using the curves in Figure 3.2.35. The saturating concentration appears to lie somewhere close to 50mM. The K_M of L-threonine is estimated to be between 7 and 9mM. It is possible that the relationship between L-threonine concentration and rate is double-sigmoidal, as with TDH from *T. brucei*. However, data covering more L-threonine concentrations would be needed to be sure of this.

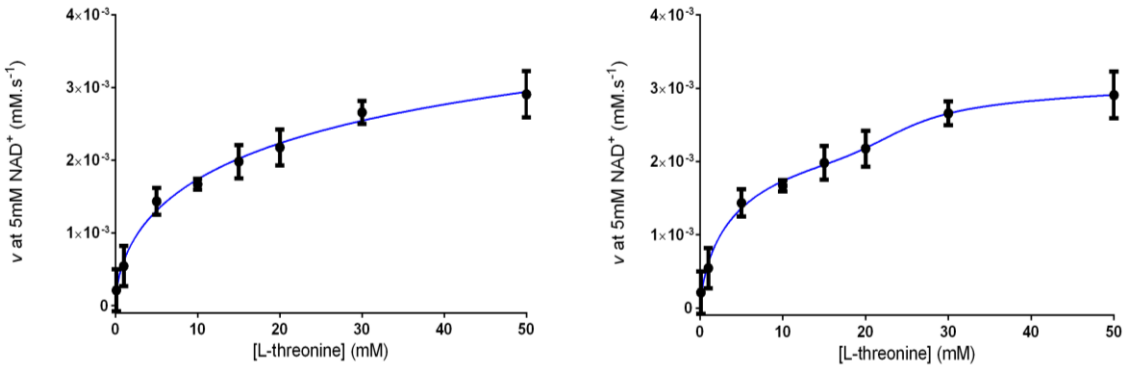


Figure 3.2.37 - plots of *Cd*TDH reaction velocity as a function of L-threonine concentration at 5mM NAD^+. The data are fit to the Hill-modification of the Michaelis-Menten equation (left) and to equation 25, which describes cooperative catalysis by two non-equivalent active sites (right).

3.3 Virtual Screening

3.3.1 Compound Libraries

3.3.1.1 Library compilation

Compilation of the compound libraries for virtual screening was carried out successfully, as described in Section 2.3.3. Please see the appendix for lists of ZINC database ID codes for the NCI diversity set II and for the ligand analogue library. In the appendix, there is also a table listing all the natural products screened, and the literature sources for their names and/or structures. The original search for proteins using Protein Structure Similarity Clustering (PSSC) found several proteins, which are listed in the appendix, along with their biological relevance, species and statistics describing their structural similarity to TDH. Ligands from the following proteins were included in the virtual screen: UDP-galactose 4-epimerase (UDP-GalE), UDP-sulfoquinovose synthase (SQD1), dTDP-glucose 4,6-dehydratase, GDP-mannose 4,6-dehydratase, CDP-tyvelose/paratose 2-epimerase, ArnA decarboxylase/dehydrogenase,GDP-4-keto-6-deoxy-D-mannose reductase, UDP-N-acetylglucosamine 4-epimerase, ADP-glycero-D-mannoheptose 6-epimerase, GDP-L-fucose synthase, CDP-D-glucose 4,6-dehydratase, GDP-mannose 3',5'-epimerase, dTDP-6-deoxy-L-lyxo-4-hexulose reductase, dihydroflavonol(dihydrokaempferol) 4-reductase (DFR), progesterone 5β-reductase and S-adenosylmethionine synthetase. The ligands from these proteins are listed in a table in the appendix. Figure 3.3.1 shows the superimposition of UDP-GalE and DFR monomers with the TDH monomer used in the PSSC search. The figure illustrates the close relationship in 3D structure, shared between different proteins matched by PSSC. All proteins had RMSDs of α-carbon atoms between 1.8 and 2.7 Å.

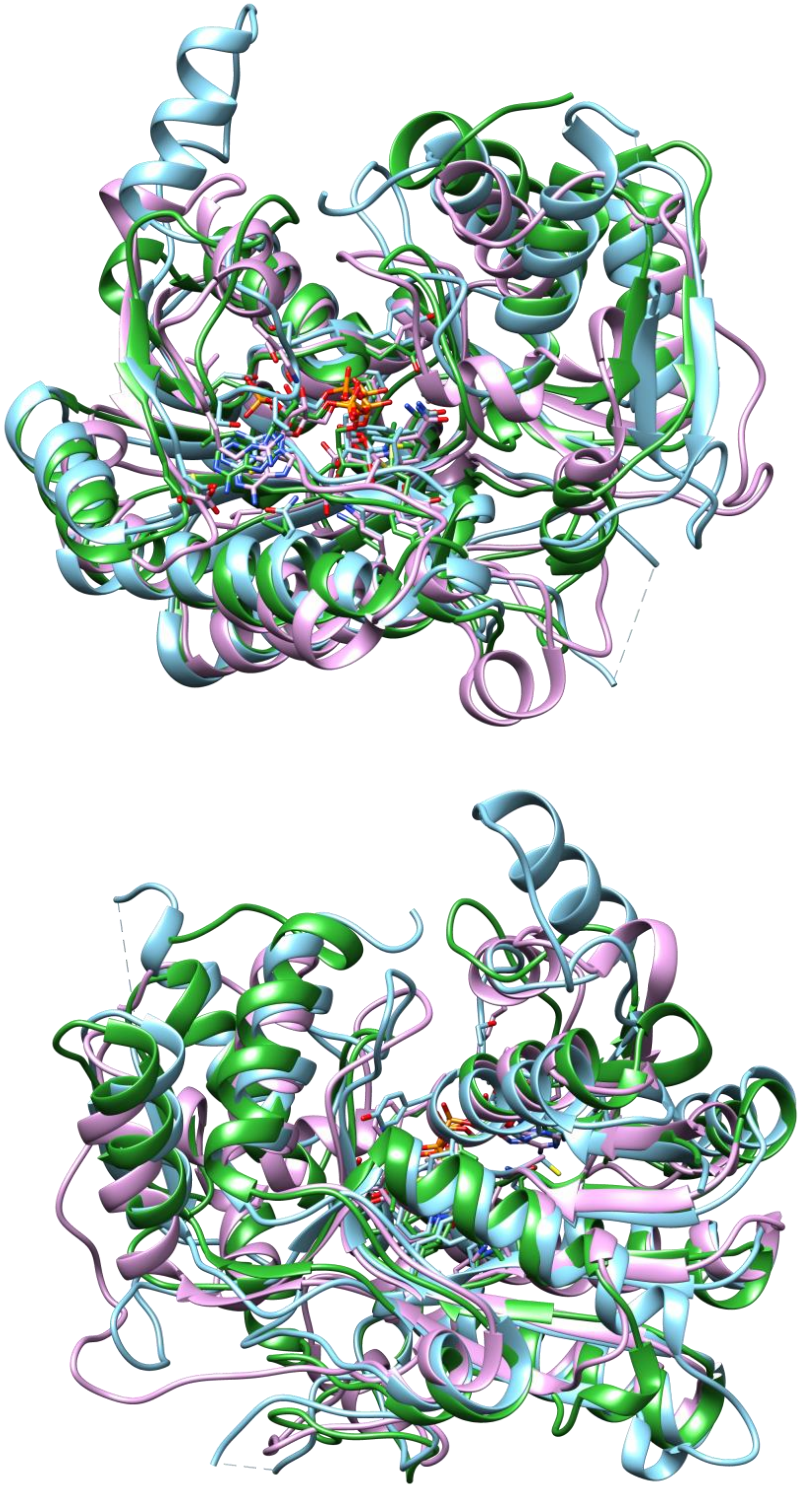


Figure 3.3.1 - ribbon representations of the structures of TDH, UDP-GalE from *Trypanosoma brucei* and DFR from *Vitis vinifera* superimposed on each other. This illustrates the degree of 3D structural similarity between the proteins, despite the fact that UDP-GalE and DFR share less than 20% sequence similarity with TDH.

3.3.1.2 Analysis of compound libraries

Figure 3.3.2 shows the visualisation of all molecules screened and displays the successful sampling of chemical space by the compound libraries (see Section 2.4.3 for a description of how the visualisation is created).

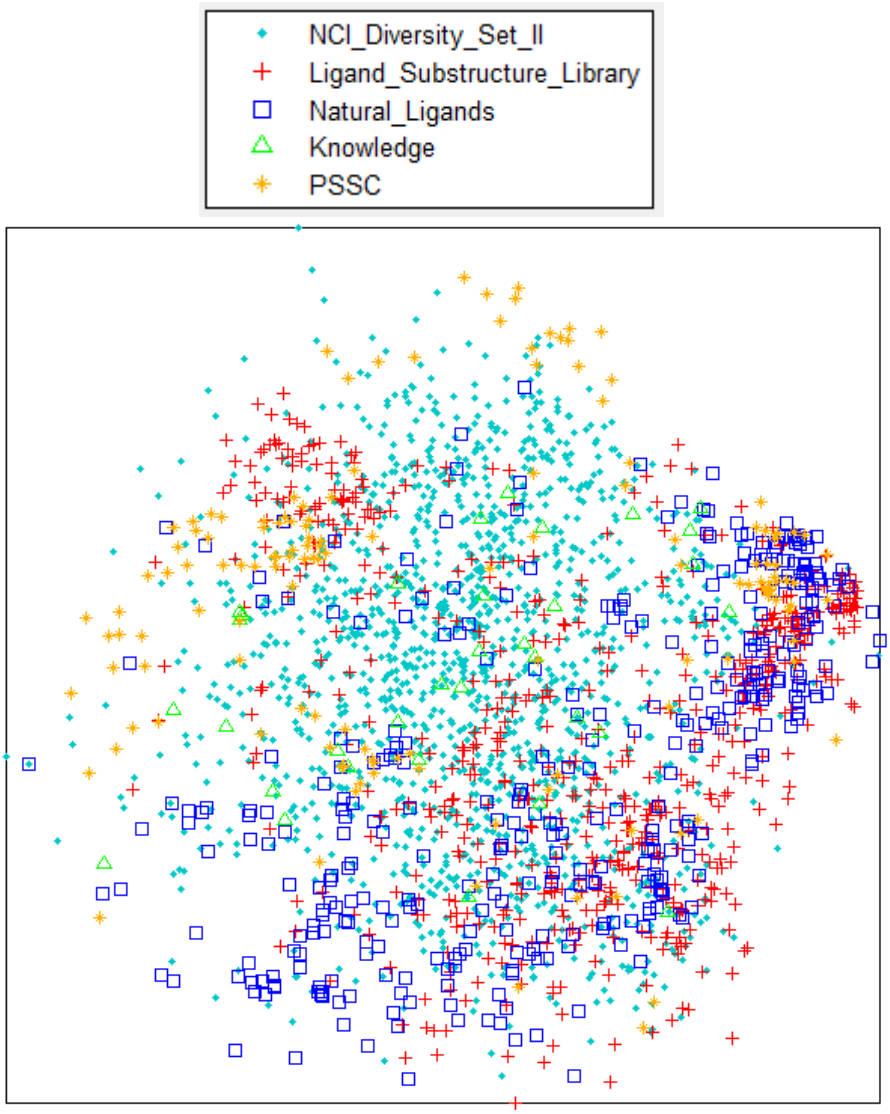


Figure 3.3.2 - visualisation of the diversity all compounds in the virtual screening library. Each point represents a single compound in the library. Its position on the diagram is determined by the difference between the compound and other compounds, as determined by the chemical fingerprint that characterises that compound.

The NCI diversity set II exhibits a wide distribution, as expected. A low mean Tanimoto similarity coefficient of 0.17 (0= no similarity, 1 = identical molecules) was calculated for the library by comparing all molecules against each other.

The natural product library appears to sample a smaller area of chemical space, and the points representing those compounds are concentrated in the lower half and right side of the data visualisation in Figure 3.3.2. In spite of this, the natural product library is still very diverse, as

represented by a Tanimoto coefficient of 0.21. The ligand analogue library appears to be even less spread across the data visualisation and there are a number of similar molecules clustered together. However, a low mean Tanimoto coefficient of 0.21 was also calculated for this library.

The 'Knowledge' and PSSC libraries have a similar distribution across the data visualisation in Figure 3.3.2; they show a wide sampling of chemical space, however clusters of similar molecules can be seen in different parts of the diagram. Although the knowledge library has a low mean Tanimoto coefficient of 0.18, there are groups of compounds that share very similar structures, notably the mouse TDH (*Mm*TDH) inhibitors which have a high mean Tanimoto coefficient of 0.87. The trypanosomiasis drugs and the compounds from the School of Pharmacy have Tanimoto coefficients of 0.15 and 0.33, respectively. The PSSC library has a low Tanimoto coefficient of 0.23. The points representing most of the molecules from the Knowledge and PSSC compound libraries are located near molecules from other compound libraries, suggesting that they would not have contributed to the overall diversity of the screening library. However, compounds from the PSSC library are represented by points near the top and left of the data visualisation – by occupying otherwise empty areas of the diagram (therefore, possessing previously unrepresented chemical structures), these compounds extend the sampling of chemical space of the entire library. Overall, the visualisation of the diversity of the compounds in the screening library shows that the design of the library was successful in achieving chemical diversity.

Although all of the compound libraries are diverse and the chemical terms listed in Table 3.3.1 will vary from molecule to molecule, clear differences can be seen when the mean chemical terms of different libraries are compared. The Lipinski rule of five describes a number of limits relating to different chemical features, which may determine a compound's likelihood of possessing good properties relating to absorption, distribution, metabolism and excretion (ADME)(277):

- No more than 5 hydrogen bond donors;
- No more than 10 hydrogen bond acceptors;
- A molecular weight (MW) of 500 Daltons or less;
- A log octanol:water partition coefficient (logP) of 5 or less.

As this is clearly important in a drug discovery context, the compliance with the rule of five is listed in Table 3.3.1. More recently, the number of rotatable bonds and a compound's polar surface area have been demonstrated to have an effect on solubility and permeability, irrespective of MW. Upper limits of 10 and 140 $Å^2$, for number of rotatable bonds and polar surface area (PSA, or topological PSA [TPSA]), respectively, have been demonstrated as predictors of oral bioavailability in drug molecules(279). A stricter guideline proposed by

Congreve and colleagues, termed the rule of three(278), has also been designed for fragment screening libraries and stipulates that it a molecule should have:

- No more than 3 hydrogen bond donors;
- No more than 3 hydrogen bond acceptors;
- A molecular weight (MW) of less than 300;
- A calculated logP (clogP) of 3 or less.

The following parameters may also be useful criteria:

- No more than 3 rotatable bonds;
- A PSA of no more than 60 $Å^2$.

These parameters and the compliance of each molecule with the rule of three (pertaining to all six criteria mentioned above) are also listed in Table 3.3.1.

Table 3.3.1 - mean chemical terms describing the compounds in different components of the virtual screening library.

	Mean Chemical Term Value									
Library	**MW**	**LogP**	**TPSA**	**Hydrogen-bond Acceptors**	**Hydrogen-bond Donors**	**Number of Rotatable Bonds**	**Lipinski Rule of Five Compliant?**	**Less than 1 Rule of Five Violation?**	**Congreve Rule of Three Compliant?**	**Mean Tanimoto**
NCI Diversity Set II	282.3	2.15	68.16	3.7	1.7	2.9	86.5% (1769 of 2044)	96.5% (1972 of 2044)	15.5% (316 of 2044)	0.17
Ligand Substructure	266.9	0.30	99.32	4.9	1.9	4.7	97.9% (734 of 750)	100% (750 of 750)	3.2% (24 of 750)	0.21
Natural	392.2	2.96	91.93	5.4	2.6	4.3	61.0% (260 of 426)	84.3% (359 of 426)	8.9% (38 of 426)	0.21
Knowledge	437.4	1.61	134.51	6.6	2.8	6.2	69.7% (23 of 33)	81.8% (27 of 33)	3.0% (1 of 33)	0.18
Trypanocidals	458.7	0.44	169.70	8.6	3.5	6.8	76.2% (16 of 21)	76.2% (16 of 21)	4.8% (1 of 21)	0.15
School of Pharmacy	313.1	2.40	77.90	3.7	1.7	4.3	100% (6 of 6)	100% (6 of 6)	0% (0 of 6)	0.33
***Mm*TDHi**	487.4	4.91	67.98	2.7	1.8	5.7	16.7% (1 of 6)	83.8% (5 of 6)	0.0% (0 of 6)	0.87
PSSC	365.5	-1.63	159.35	7.9	4.3	5.5	57.2% (99 of 173)	65.9% (114 of 173)	6.4% (11 of 173)	0.23
All	299.0	1.65	83.34	4.4	2.0	3.6	82.0% (2808 of 3426)	91.6% (3138 of 3426)	11.2% (383 of 3426)	

The majority of the compounds screened were Rule of Five compliant, though some libraries were more compliant than others. Only 61.0% and 69.7% of the molecules in the Natural product and Knowledge libraries, respectively, were Rule of Five compliant, compared to 82.0% of the compounds in the screening library as a whole. This is probably due to the higher molecular weights of the compounds in these two libraries. The PSSC library is the least compliant with the Lipinski rule of five – only 52.2% of the molecules screened complied with the rule and only an additional 8.7% had one violation. The mean numbers of hydrogen bond donors and acceptors is the highest for this library, so it is likely that several of the molecules violate the rule of five in this regard. As none of these libraries were designed to be fragment libraries, it is unsurprising that very few compounds in the screening library are compliant with the rule of three.

3.3.2 Assay Validation

Four different TDH structure models were used as receptors in the four screens: Virtual Screens 1, 2, 3 and 4. The model for Virtual Screen 1 was the monomer in subunit F bound to NAD^+ in the structure denoted TK8. The model for Virtual Screen 2 was a monomer (subunit C) and the bound NAD^+ in the model TIO8, which had flexible loop 1 in a closed position over the binding site. For Virtual Screen 3, the structure was taken from the model denoted TS2, which contained an *apo*-TDH structure, with the flexible loop in the open conformation. Finally, for Virtual Screen 4, the same model as for Virtual Screen 2 was used, but the bound NAD^+ molecule was removed from the PDB file.

As described in the methods (Section 2.3.2), five active site residues that were identified as being important in L-threonine binding (Ser82, Thr119, Met81, Tyr144 and Thr186), were chosen as variables, where they were set as rigid or flexible, to determine their effect on AutoDock docking predictions. A number of docking simulations were carried out using default Lamarckian Genetic Algorithm parameters with different combinations of flexible residues and a completely rigid TDH, using threonine, glycerol or pyruvate as a ligand. The validity of the rigid or flexible 'receptor' model was judged based on the plausibility of the binding position for L-threonine, the predicted binding energies and the clustering of binding poses. A completely rigid molecule was chosen, partly due to the more accurate binding energies predicted for all ligands. Flexible models produced energies corresponding to K_i values in the micromolar and nanomolar range, compared to values of 1.42mM for L-threonine and 2.09mM for pyruvate which are within up to approximately one order of magnitude of published L-threonine K_M values(139) and the K_i value of pyruvate determined in this lab, 12.7mM (note that a K_i of 17.6mM was determined in this thesis). Another reason for the choice of the rigid TDH model was the predicted binding poses of the ligands. For example, when testing the proposed

models with glycerol, the poses that were most similar to those found in the TK8 model were produced by a rigid structure. These binding poses were also repeatable and all ten AutoDock runs predicted poses within an RMSD of 2 Å from each other. Conversely, the models with flexible residues often predicted widely different binding poses. Although the rigid TDH models were ultimately selected for all TDH screens, models where the Cα→Cβ bonds in the side chains of Ser82 and/or Tyr144 were rotatable were able to predict ligand binding poses as accurately as the rigid model.

The decision to screen rigid TDH models was largely based on the results of tests on variations of the model for Virtual Screen 1. However, for subsequent TDH screens (2-4), validation studies were carried out to confirm the validity of the screening results. Docking experiments with L-threonine against the model used for TDH screen 2 (which resembled the TDH conformation in the TNADH4 model, see above) gave binding poses that were very similar to those found in an X-ray crystallographic structure (TNADH4, see Table 3.1.1). Figure 3.3.3 demonstrates how a structural alignment of the TNADH4 structure with the docked L-threonine-TDH prediction shows that the binding prediction is almost identical to the observed binding position.

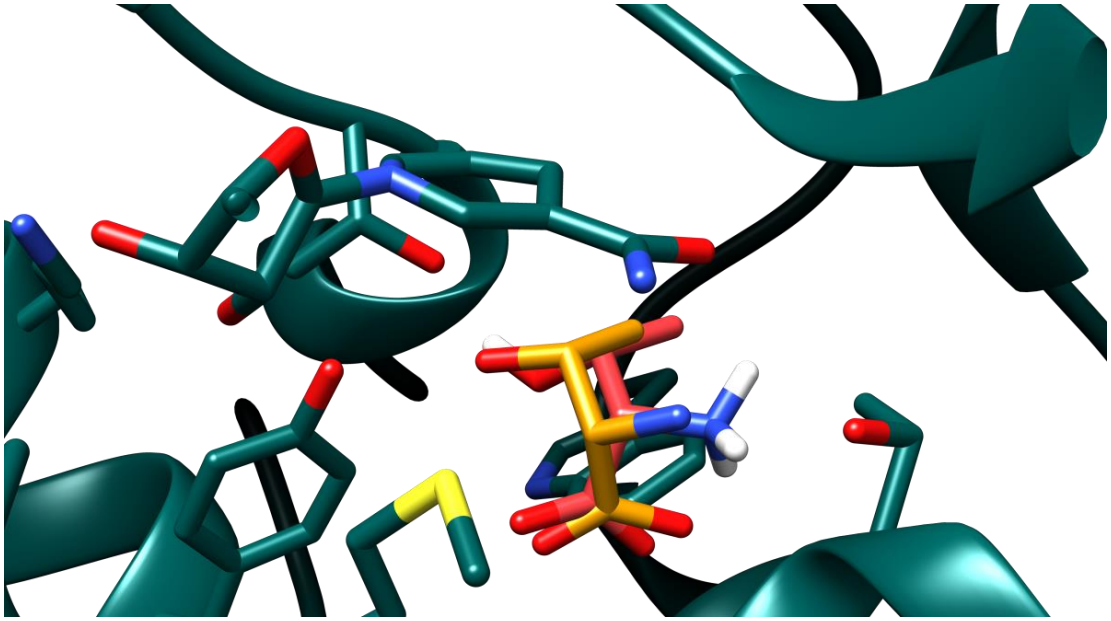


Figure 3.3.3 - a superimposition of an L-threonine-bound TDH structure (obtained by X-ray crystallography) on a Virtual Screening model with its predicted binding mode for L-threonine. The crystallographic TDH model is coloured green, and the bound TDH model is coloured gold. The predicted binding pose of L-threonine is depicted in pink and the model used in screening (in Virtual Screen 2) is removed for clarity. It can be seen that the two binding positions are almost identical.

As information on the binding position of NAD^+/NADH was also available, more validation experiments were carried out using NAD^+ as a ligand.

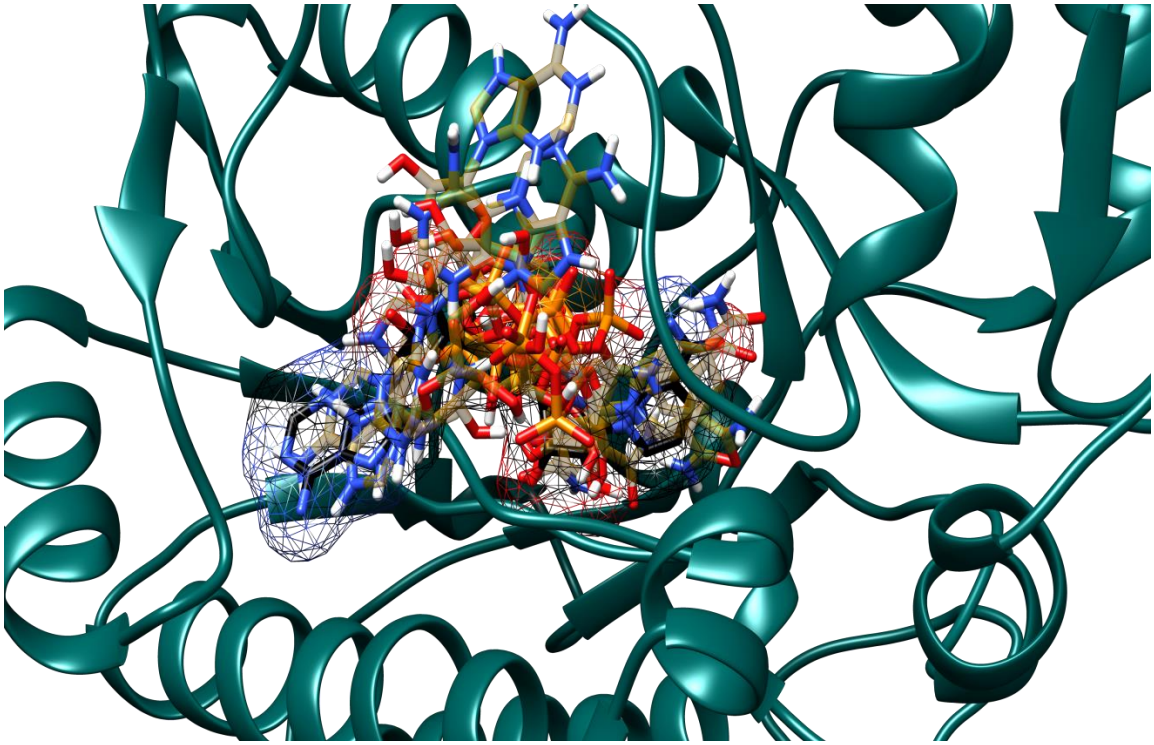


Figure 3.3.4 - predicted binding poses of NAD^+ compared to the binding pose found by virtual screening. TDH is coloured green and the bound NAD^+ is coloured black and surrounded by a wire mesh to highlight the binding position. Predicted binding poses of NAD^+ are coloured light orange. It can be seen that Autodock was unable to fully predict the binding pose of NAD^+.

As can be seen in Figure 3.3.4, the accurate prediction of the NAD^+ binding position was not so successful. This is actually expected, due to the large number of rotatable bonds in the molecule and the limitation in AutoDock of accurately predicting the binding of ligands with more than 10 rotatable bonds(24).

3.3.3 Hits

There were huge differences in the hit rates of each TDH screen. At a binding energy cut-off of -9.5 $kcal.mol^{-1}$, the first Virtual Screen produced a modest hit rate of 2.2%, whilst Virtual Screen 2 produced no hits at all. Virtual Screens 3 and 4, however, produced much higher hit rates of 7.3% and 8.5% respectively.

Table 3.3.2- hit rates achieved in Virtual Screens 1, 3 and 4.

	Hit Rate		
Library	Virtual Screen 1	Virtual Screen 3	Virtual Screen 4
NCI Diversity Set II (n = 2044)	2.7%	5.3%	5.9%
Ligand Substructure (n = 750)	0.9%	0.5%	2.5%
Natural (n = 426)	3.1%	27.9%	29.6%
Knowledge (n = 33)	3.0%	12.1%	15.2%
Trypanocidals (n = 21)	0.0%	0.0%	4.8%
School of Pharmacy (n = 6)	0.0%	33.3%	16.7%
***Mm*TDHi (n = 6)**	0.0%	33.3%	50.0%
PSSC (n = 173)	0.6%	8.1%	12.1%
All (n = 3046)	2.2%	7.3%	8.5%

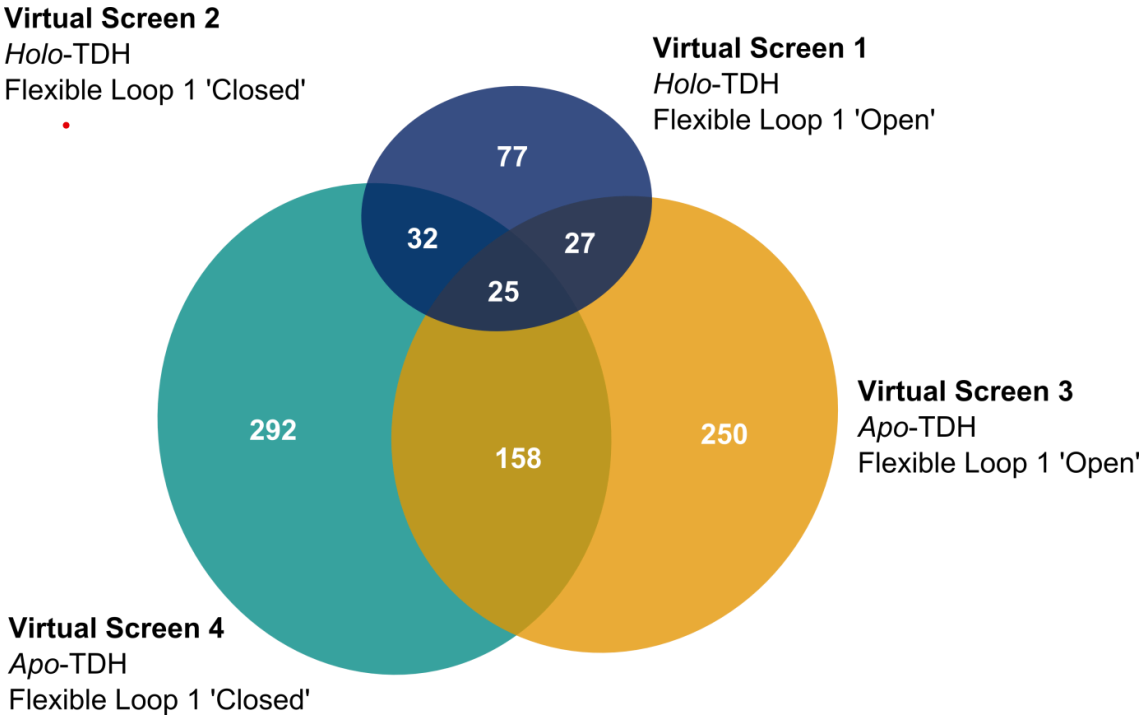


Figure 3.3.5 - Venn diagram outlining the number of hits in each screen and the common hits between screens. The diagram was constructed using the software eulerAPE (280).

In Virtual Screen 1, the NCI diversity set II, the Natural Product library and the Knowledge library all produced similar hit rates. It is also evident that the NCI diversity set II had comparable hit rates across all successful screens. In Virtual Screens 3 and 4, however, the natural product library produced huge hit rates of 27.9% and 27.6%, respectively. The hit rates of the Knowledge and PSSC libraries were also greatly increased, although to a lesser extent than the natural product library. These large differences in hit rate can most likely be attributed to the availability of the NAD^+ binding site. Referring back to Table 3.3.1, it can be seen that the average molecular weights of the natural product, Knowledge and PSSC libraries are

significantly higher than those of the other libraries. Thus, with the absence of NAD^+ from its binding pocket in TDH screens 3 and 4, the larger molecules have more space to bind and make stronger interactions with the protein. The total lack of hits in the second virtual screen can be explained by the fact that NAD^+ was present, whilst flexible loop 1 was in its closed conformation, leaving very little room for binding in the substrate binding sites. Only the smallest molecules could be docked into the L-threonine binding pocket, whilst others bound to the exterior surface of TDH. For this reason, the top scoring compound had a binding energy of -7.6kcal.mol^{-1}, which is higher (or weaker) than the cut off of -9.5kcal.mol^{-1}.

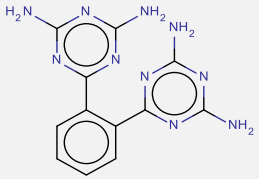
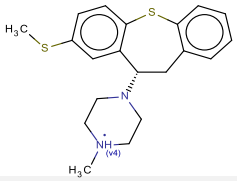
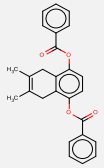
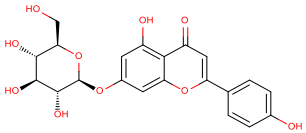
Table 3.3.3 – mean chemical terms describing the features of the hits from the three successful Virtual Screens.

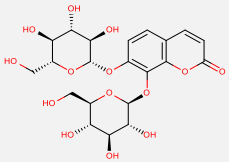
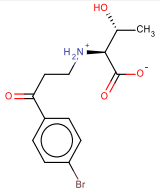
	Mean Chemical Term Value									
Library	**MW**	**LogP**	**TPSA**	**Hydrogen-bond Acceptors**	**Hydrogen-bond Donors**	**Number of Rotatable Bonds**	**Lipinski Rule of Five Compliant?**	**Less than 1 Violation?**	**Congreve Rule of 3 Compliant ?**	**Hit Rate**
Virtual Screen 1	362.18	1.34	91.62	5.4	2.6	3.5	72.7% (56 of 77)	90.9% (70 of 77)	2.6% (2 of 77)	2.20%
Virtual Screen 3	451.41	3.70	102.04	5.9	2.7	4.3	39.6% (99 of 250)	78.0% (195 of 250)	0.0% (0 of 250)	7.30%
Virtual Screen 4	440.32	3.45	102.77	5.8	2.8	4.4	42.5% (124 of 292)	82.2% (240 of 292)	0.0% (0 of 292)	8.52%

The above observations are emphasised by the results in Table 3.3.3, above. The mean molecular weights of the compounds in Virtual Screens 3 and 4 are higher than in Virtual Screen 1. It is interesting that all of the mean molecular weights of the hits are higher than the mean molecular weight for all the compounds in the screening library. This suggests that larger compounds were more likely to be predicted to bind more tightly. Despite the higher molecular weights of the hits in Virtual Screens 3 and 4, the mean numbers of hydrogen bond donors and acceptors, and the number of rotatable bonds are similar across all screens. The majority of the hits have one violation or less of the Lipinski rule of five, though the hits in screen 1 are more compliant with the rule. This is probably due to the higher molecular weight and LogP of the hits in the other screens.

Table 3.3.4 lists the top 5 hits and other hit compounds of interest, including their names and structures, from Virtual Screen 1, which had a hit rate of 2.2%.

Table 3.3.4 – the top 5 Hits from Virtual Screen 1, and one hit that later exhibited activity *in vitro* (see Section 3.4)

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors, Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-11.88, 1.96	ZINC01673377		296.29	1.72	181.42	10 , 4	2	✓		
-11.6, 3.14	ZINC19362650		357.56	0.73	7.68	2 , 1	2	✓		
-11.43, 4.19	ZINC01755448		398.45	6.87	52.6	2 , 0	6		✓	✓
-11.42, 4.26	apigenin-7-glucoside		432.38	0.44	166.14	10 , 6	4		✓	✓

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors, Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-11.34, 4.87	daphneside		502.42	-3.36	225.06	13 , 8	6			✓
-9.51, 106.92	ZINC04141542 (BPOB)		330.17	-1.01	94.04	4 , 2	7	✓	✓	

The diversity of the hits from the first virtual screen is visualised in Figure 3.3.6. Comparing this to the visualisation of the entire screening library in Figure 3.3.2, there is no indication that the hits constitute groups of very similar molecules.

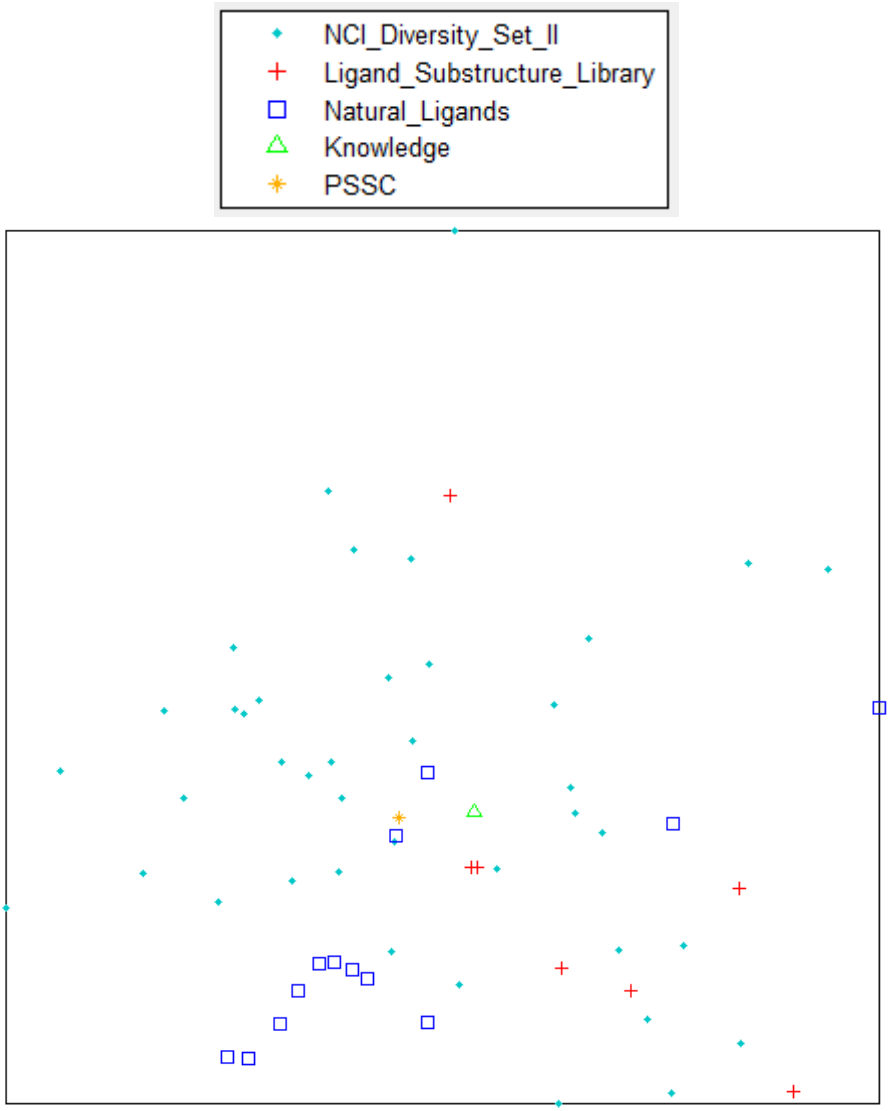


Figure 3.3.6 - visualisation of the hits from Virtual Screen 1. Each point represents a hit and the relative positions of the points are determined by the differences between their structures, as in Figure 3.3.2.

There does appear to be more sampling of the lower half of the diagram, but this portion was more populated to begin with. Less of the compounds represented by points in the upper half of the diagram are hits than one would expect from a random selection from the screening library, so it is possible that compounds with certain structural features did not score highly in this screen.

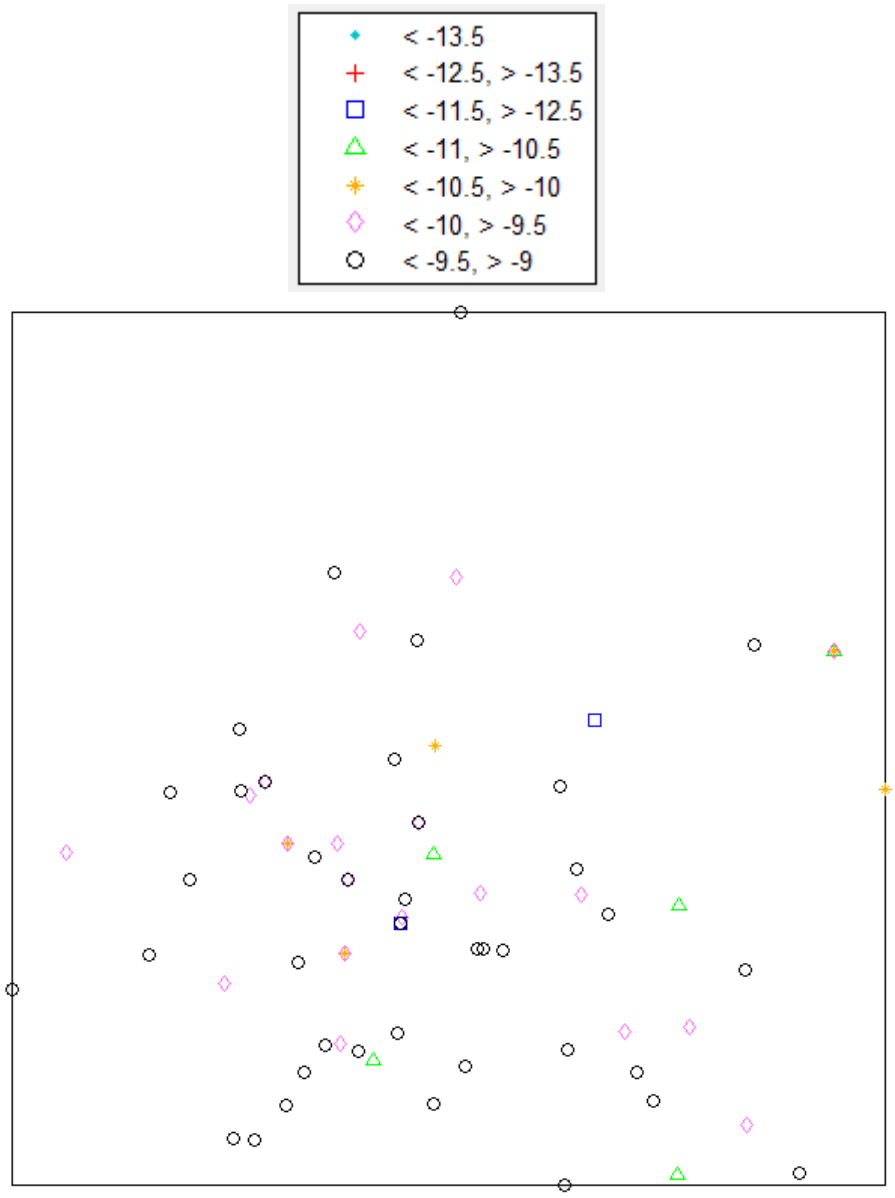
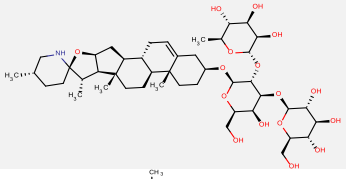
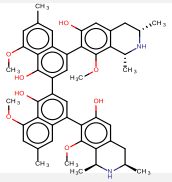
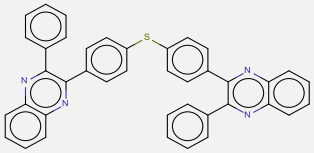
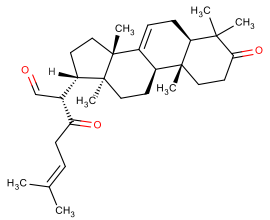


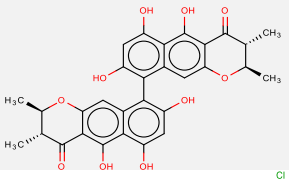
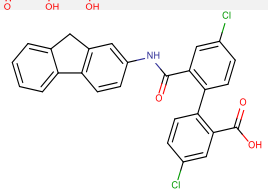
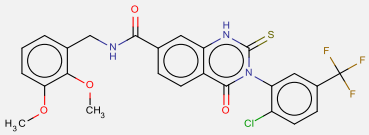
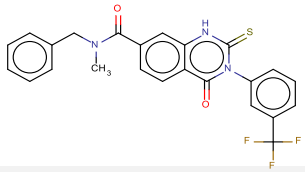
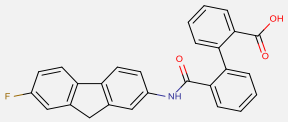
Figure 3.3.7 - visualisation of the diversity of hits from Virtual Screen 1. The compounds are represented by different shapes, according to their predicted binding energies.

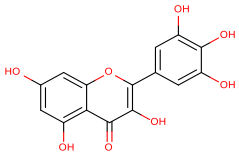
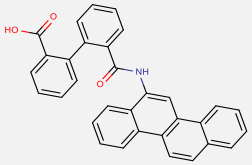
Looking at Figure 3.3.7, there is no clear relationship between the structures of hits, as represented by a point on the data visualisation, and their predicted binding energies. The majority of the hits are in the lower binding energy ranges.

The following table lists the top 5 hits and other hit compounds of interest from Virtual Screen 3.

Table 3.3.5 – the top 5 Hits from Virtual Screen 3 and hits which either exhibited activity *in vitro* or an analogue exhibited activity *in vitro* (see Section 3.4)

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-15.03, .01	α-solamargine (solasonine)		884.06	0.34	258.71	17 , 10	8			
-13.62, .10	ancistrogriffithine A		784.94	7.21	141.9	10 , 6	5			
-13.62, .10	ZINC05492794		594.73	10.95	51.56	4 , 0	6			✓
-13.11, .25	3,23-dioxotirucalla-7,24-dien-21-al		452.67	6.42	51.21	3 , 0	5			✓

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-12.72, .47	ZINC17465983		546.52	5.67	173.98	10 , 6	0			✓
-11.48, 3.85	NSC132249 (CHEMBL1173812)		474.34	7.34	66.4	3 , 2	4		✓	
-10.45, 21.88	Qc2		549.95	5.48	79.9	4 , 2	6			✓
-9.96, 50.03	Qc6		469.48	5.42	52.65	2 , 1	5			
-9.94, 51.75	NSC128598 (CHEMBL1173453)		423.44	6.28	66.4	3 , 2	4			

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.8, 65.54	myricetin		318.24	1.85	147.68	8 , 6	1			✓
-9.57, 96.63	NSC128608 (CHEMBL1173811)		467.51	7.34	66.4	3 , 2	4			✓

Virtual Screen 3 had a much larger hit rate of 7.3%. In further contrast with Virtual Screen 1, one can see that there are a number of very high affinity molecules at the top of Table 3.3.5. These molecules are of high molecular weight, and α-solamargine and ancistrogriffithine A, in particular, possess a large number of hydrogen bond donors and acceptors, so they are able to make a large number of interactions with the protein. Another observation of the hits from this screen is that there are many more natural products represented in Table 3.3.5 than in Table 3.3.4.

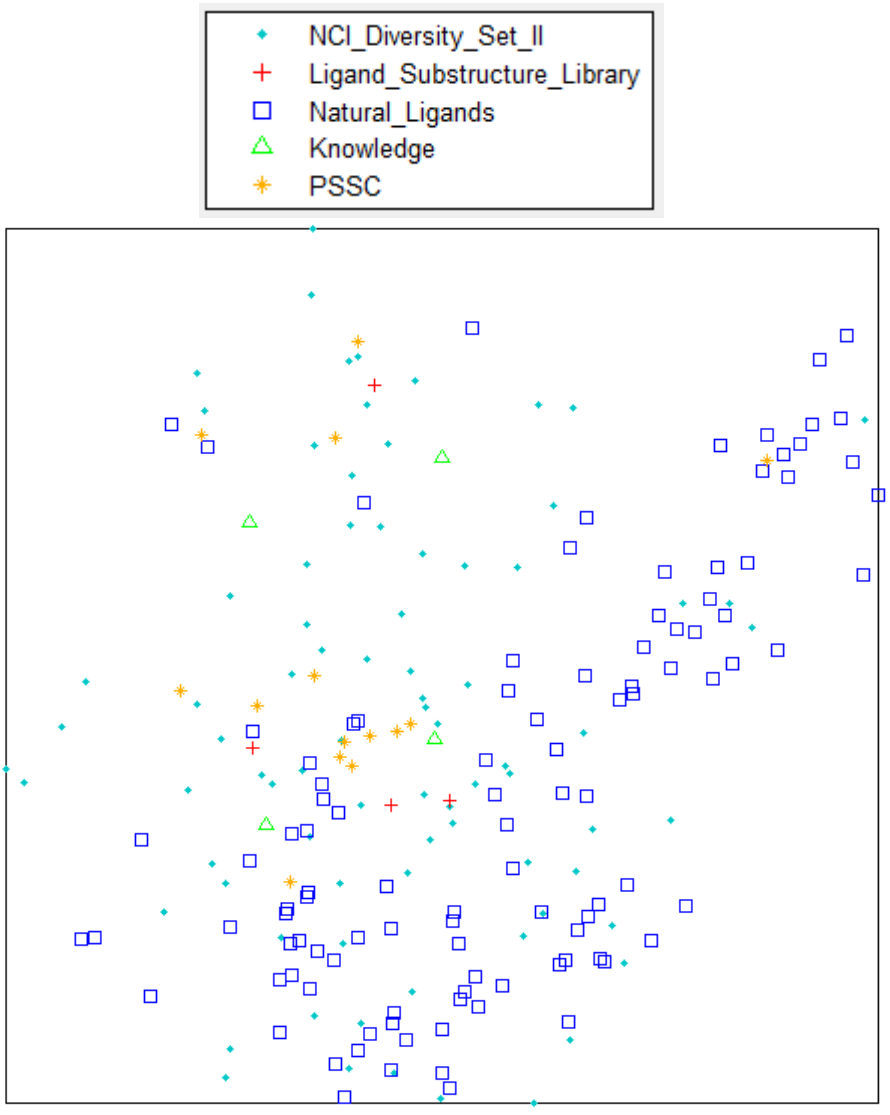


Figure 3.3.8 - visualisation of the diversity of the hits from Virtual Screen 3. Each point represents a different compound and the relative positions of the points are determined by the differences between their structures, as in Figure 3.3.2.

As with the visualisation of the chemical diversity of the Virtual Screen 1 hits, there is no evidence of hits being clustered in certain areas of chemical space in this screen. In fact, the hits in Virtual Screen 3 appear to be more diverse than those of the first screen. There is also more sampling of space in the upper portion of the diagram.

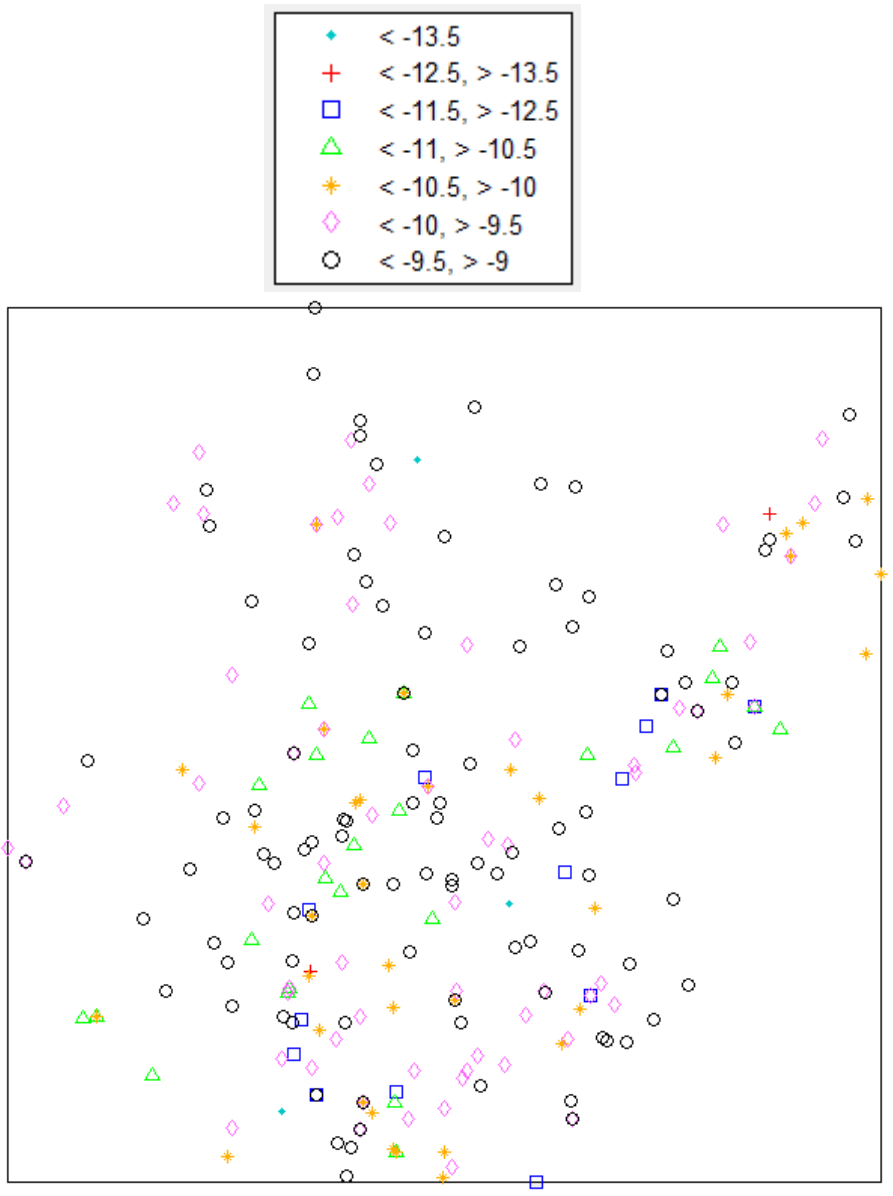
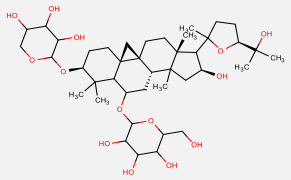
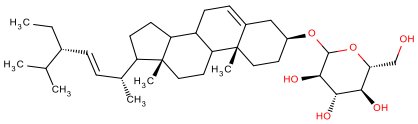
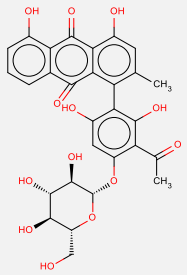
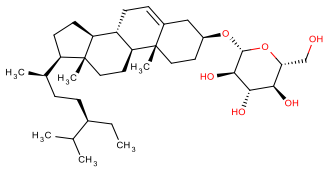


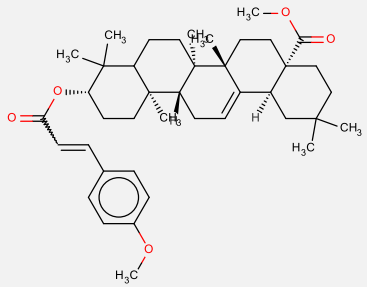
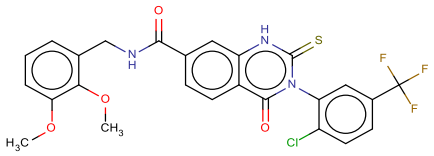
Figure 3.3.9 - visualisation of the diversity of the hits from Virtual Screen 3. Each point represents a single compound. The compounds are represented by different shapes, depending on their predicted binding energies.

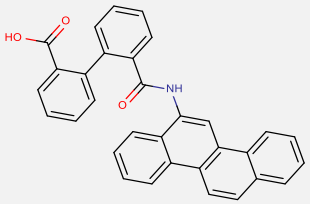
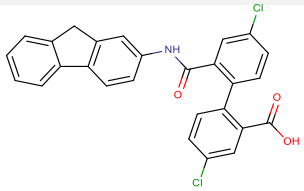
The visualisation in Figure 3.3.9 suggests that there is no relationship between structure and the binding affinity predicted by AutoDock4. The presence of more hits with higher binding affinities than in the first screen is evident.

Table 3.3.6 lists the tip 5 hits and other hit compounds of interest from Virtual Screen 4.

Table 3.3.6 – the top 5 Hits from Virtual Screen 4 and hits which either exhibited *in vitro* activity or have an analogue which exhibited activity *in vitro* (see Section 3.4).

Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-14.88, .01	astrogaloside III		784.97	-0.31	228.22	14 , 9	7			✓
-13.2, .21	stigmasterol-D-glucoside		574.83	5.71	99.38	6 , 4	8			✓
-12.62, .56	4'-O-demethylknipholone-4'-O-β-D-glucopyranoside		582.51	2.80	231.51	13 , 8	4			✓
-12.48, .71	β-sitosterol glucoside		576.85	6.07	99.38	6 , 4	9			✓

Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-12.39, .83	*cis-p*-methoxycinnamoyloxyoleanolic acid methyl ester		630.90	9.62	61.83	3 , 0	7			✓
-10.81, 11.92	Qc2		549.95	5.48	79.9	4 , 2	6			✓

Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.31, 27.71	NSC128608 (CHEMBL1173811)		467.51	7.34	66.4	3 , 2	4			✓
-10.23, 31.72	NSC132249 (CHEMBL1173812)		474.34	7.34	66.4	3 , 2	4			✓

The fourth virtual screen had the highest hit rate, at 8.5%. The results of this screen share many similarities with the previous screen. There are a large number of natural products which were hits and there are some high molecular weight compounds with very strong binding affinities. The natural product astrogaloside III, for example, was predicted to bind with a K_i of 10pM and has 14 hydrogen bond acceptors and 9 hydrogen bond donors.

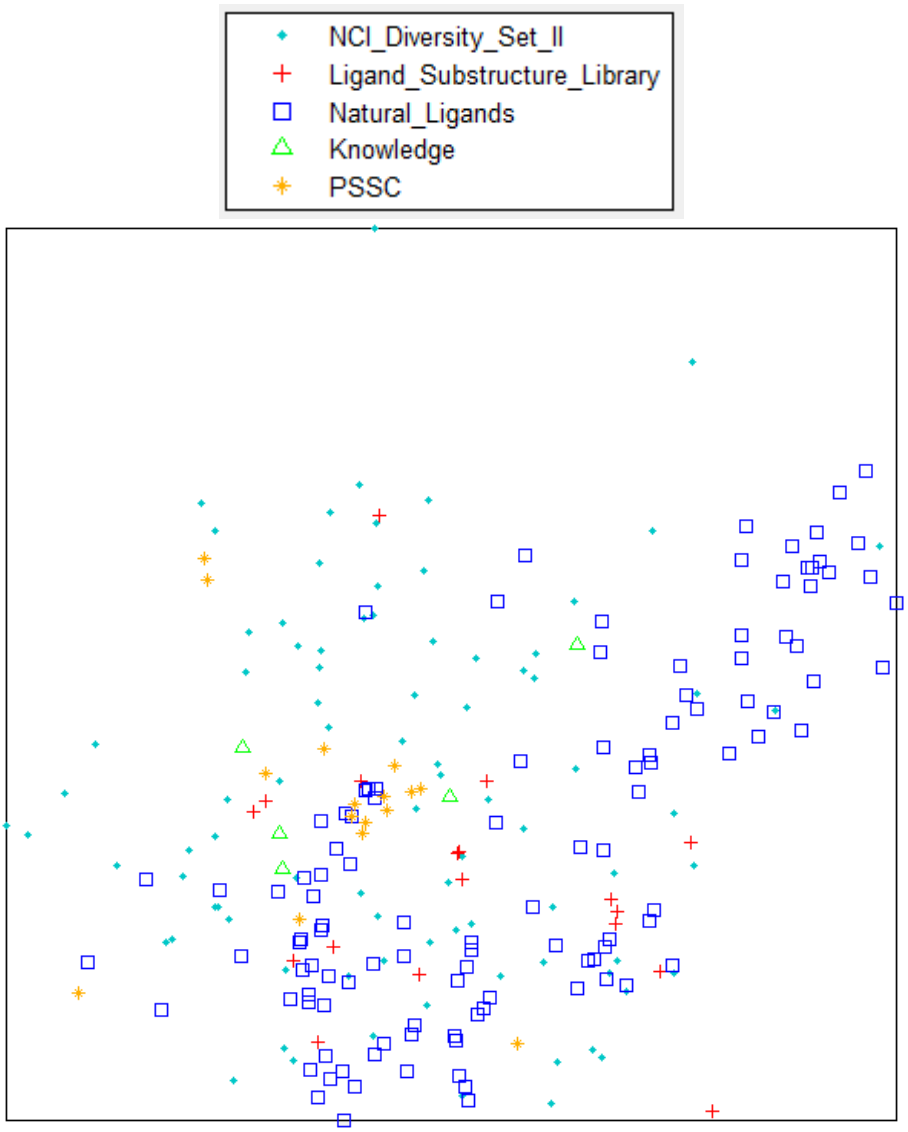


Figure 3.3.10 - visualisation of the diversity of the hits from Virtual Screen 4. Each point represents a single compound and the relative positions between the points depend on the differences between the structures of the compounds they represent, as in Figure 3.3.2.

The visualisation of the Virtual Screen 4 hits in Figure 3.3.10 is very similar to that of the Virtual Screen 3 hits in Figure 3.3.8. Many of the same compounds seem to be represented in both diagrams and the distributions of the compounds are quite similar. One of the main differences is that there are several more ligand analogues represented in Figure 3.3.10. This is due to

their higher hit rate: this library had a hit rate of 2.5% Virtual Screen 4, compared to a hit rate of 0.9% and 0.5% in Virtual Screens 1 and 3, respectively.

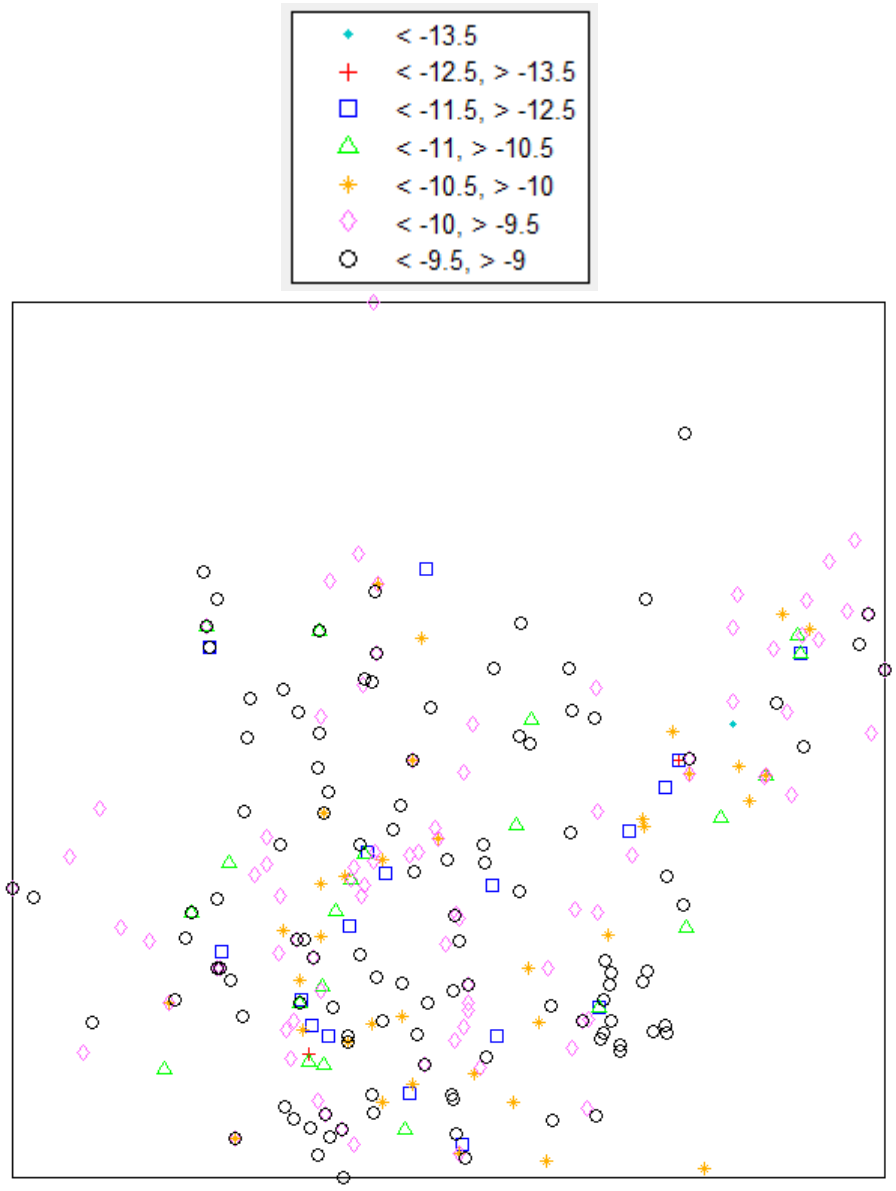


Figure 3.3.11 - visualisation of the diversity of hits from Virtual Screen 4. Each point represents a single compound. Each compound is represented with a different shape, depending on its predicted binding energy.

As with the other screens, there is no clear evidence of a relationship between the molecular structures of the hits and their predicted binding energies. Comparing all the visualisations of hit diversity allows a comparison of each of the screens. Virtual Screen 1 has the smallest sampling of chemical space and its visualisations are less populated. As the area available for binding on the TDH model used in this screen was smaller, the compounds were able to make less interactions with the target and had lower binding energies. Virtual Screen 3 hits seem to show the largest sampling of chemical space as the TDH model used in this screen had the largest binding site. Virtual Screen 4 used a TDH model which also had no NAD bound, but

with flexible loop 1 in the closed conformation. The hits had slightly less sampling of chemical space than in the third screen, but the overall hit rate was higher.

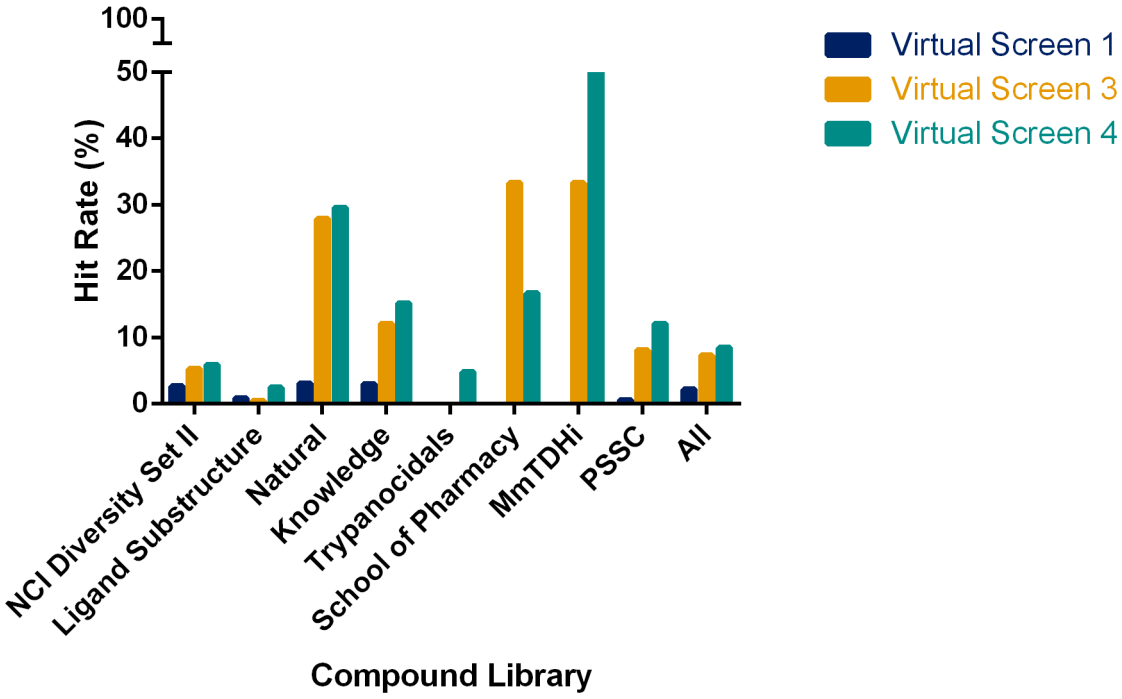


Figure 3.3.12 - column chart representation of virtual screening hit rates. Hit rates are grouped according to compound library.

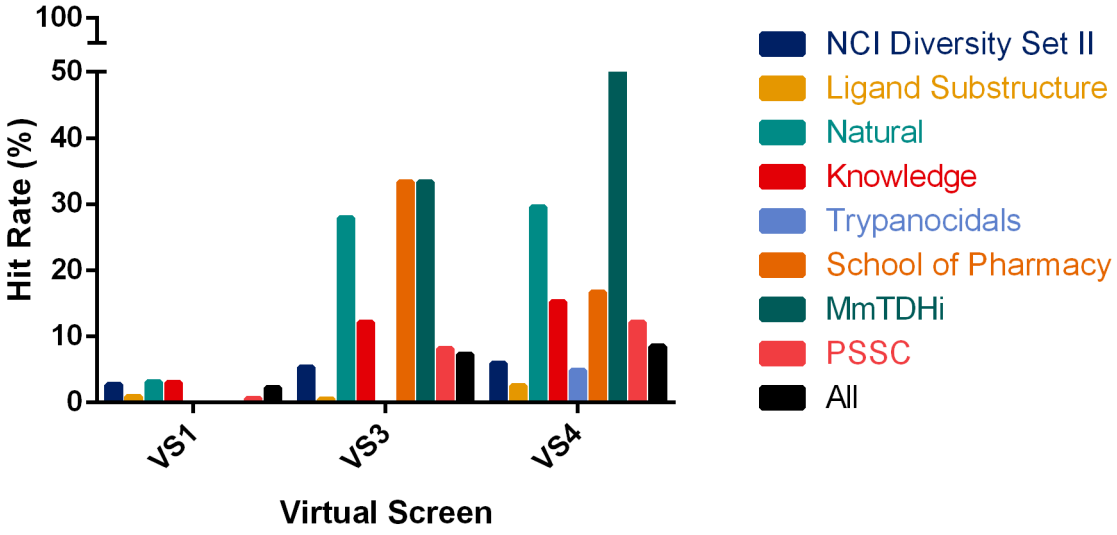


Figure 3.3.13 - column chart representation of virtual screening hit rates. Hit rates are grouped according to Virtual Screen.

The following section examines the binding modes of the strongest hits in each screen and highlights the relationships between the predicted binding modes and the TDH conformation used in each screen.

3.3.4 Binding Interactions

3.3.4.1 Virtual Screen 1

BPOB

The L-threonine analogue 2-{[3-(4-bromophenyl)-3-oxopropyl]amino}-3-hydroxybutanoic acid (BPOB or ZINC04141542) was predicted to bind TDH with a K_i of 107nM in Virtual Screen 1. An examination of the predicted binding poses of this compound helps to highlight the important binding interactions made by hits in Virtual Screen 1.

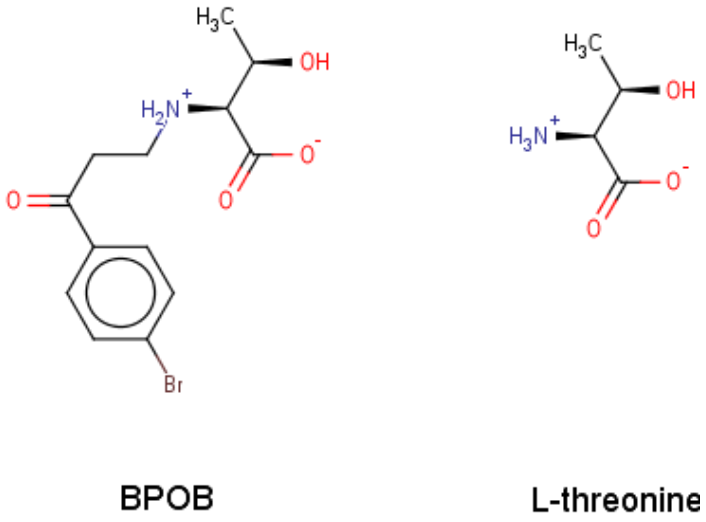


Figure 3.3.14 - the molecular structures of BPOB and L-threonine

AutoDock predicted two distinct binding modes: either with the bromobenzene group pointing into the L-threonine binding pocket, or with the bromobenzene group pointing outwards. In both binding positions, the L-threonine binding pocket was occupied, and BPOB formed interactions with the same residues as L-threonine: Ser82, Thr119, Tyr144, Thr186 and Trp280. It also interacts with other nearby residues and with the phosphate group of NAD^+.

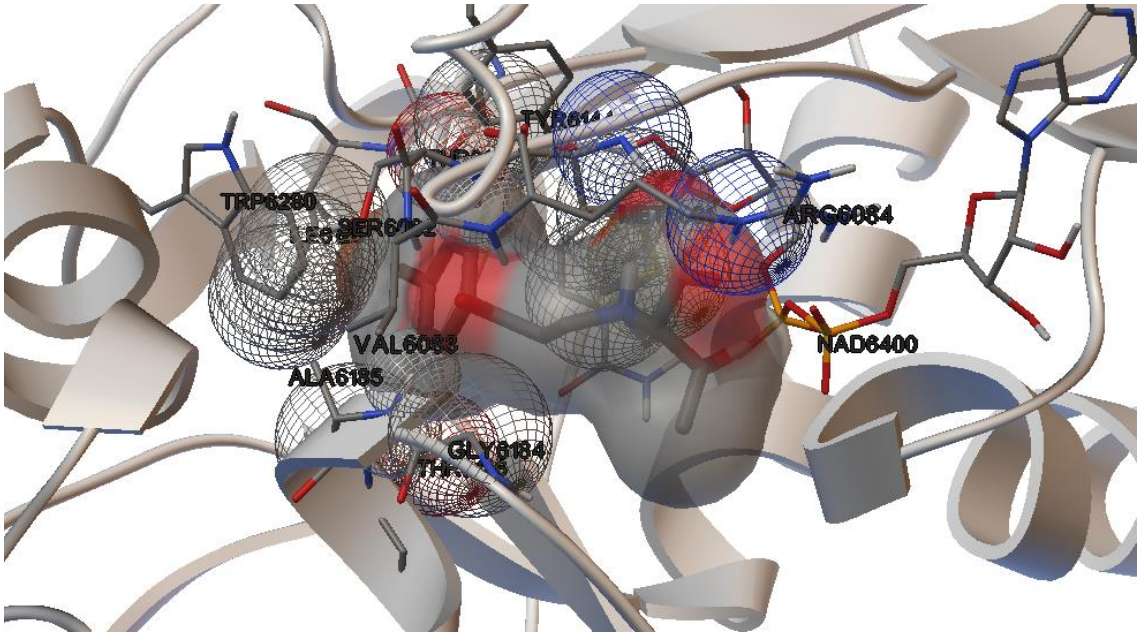


Figure 3.3.15 - a predicted binding mode of BPOB, predicted by AutoDock in Virtual Screen 1. BPOB is surrounded by a semi-transparent molecular surface and is positioned with the bromine atom facing the interior of the protein. Wire spheres are shown around TDH atoms that are interacting with the ligand.

The two predicted binding modes are shown next to each other in Figure 3.3.16. In the lower image, the L-threonine-like part of the molecule occupies the L-threonine binding pocket in a similar fashion to L-threonine itself (see Figure 3.1.39, in section 3.1), with the β-carbon of the L-threonine moiety lying close to the C4 atom of the NAD^+ nicotinamide group.

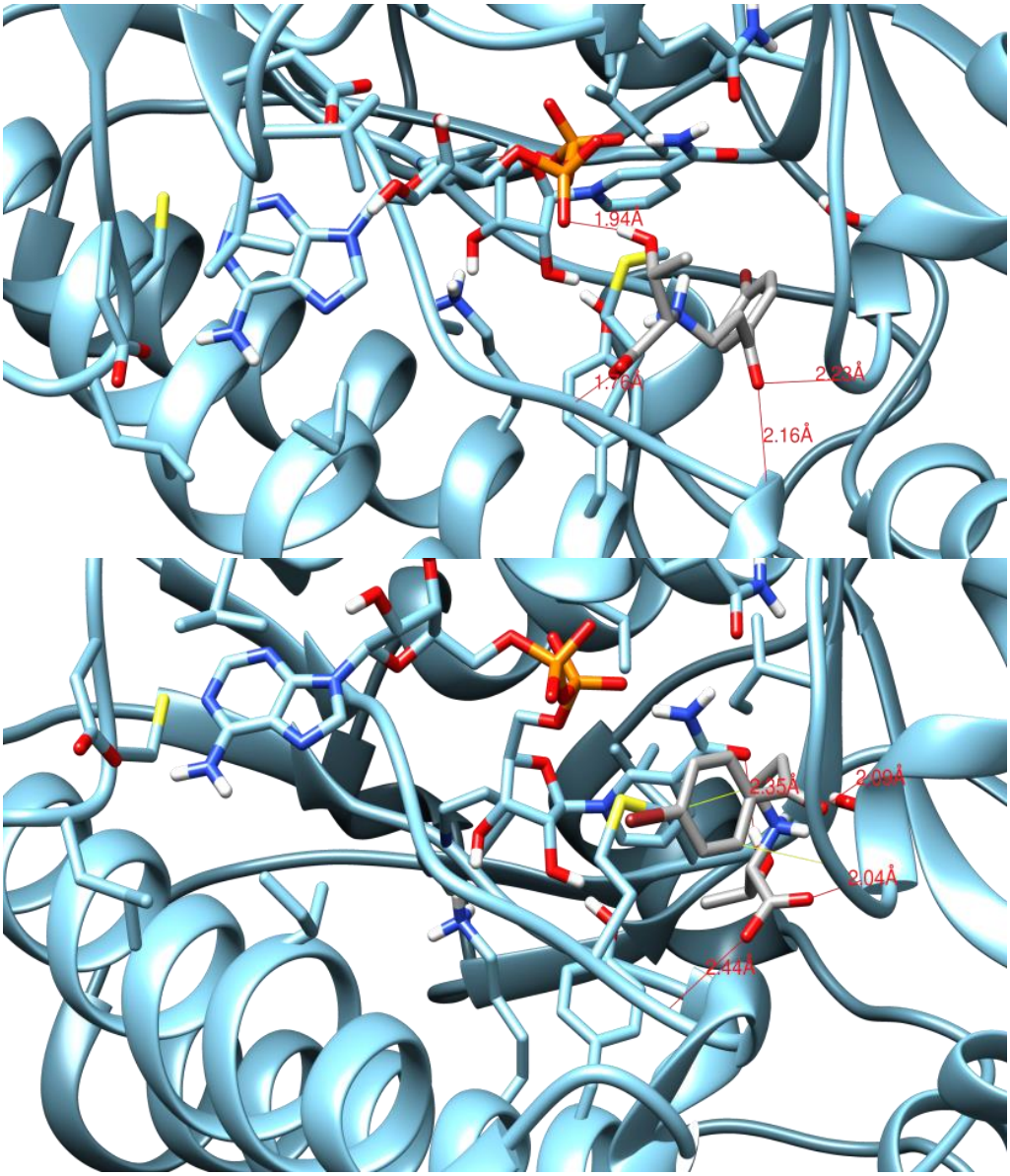


Figure 3.3.16 - the two predicted binding modes of BPOB in Virtual Screen 1. The bromobenzene group binds inside the L-threonine binding pocket (top image) or points outwards from the active site (bottom image). Hydrogen bonds are shown by red lines and labelled with the bonding distances

Similar interactions with TDH can be seen as one examines the binding modes of the other top hits.

ZINC1673377

The top scoring compound in Virtual Screen 1 was ZINC1673377, which was predicted to bind with a K_i of 1.96nM.

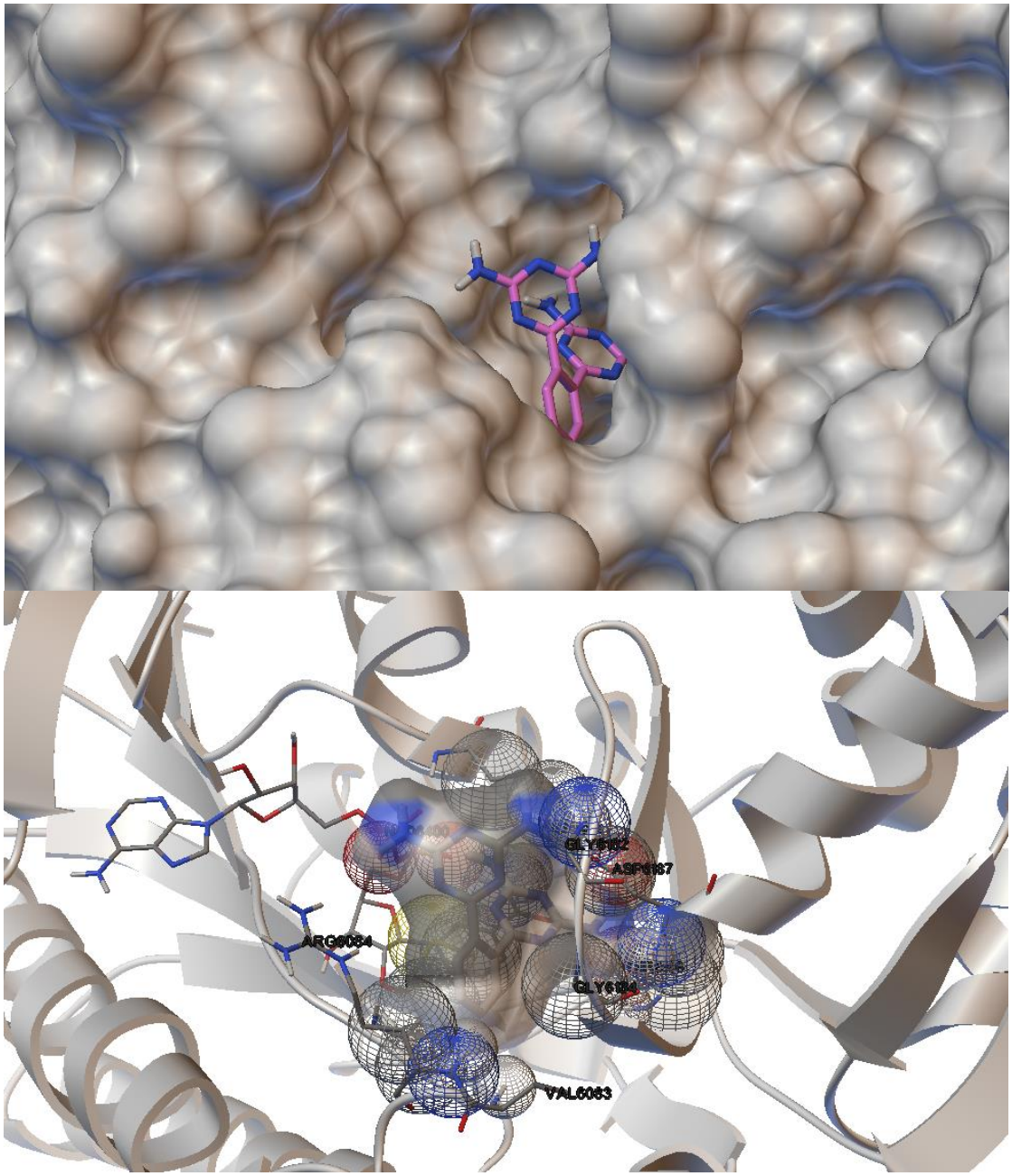


Figure 3.3.17 - predicted binding mode of ZINC1673377 depicted in AutoDock Tools (ADT). The top image shows ZINC1673377 bound to a molecular surface representation of TDH. The lower image highlights the binding interactions that it makes with TDH.

ZINC1673377 makes interactions with the important L-threonine-binding residues Thr186 and Met 81. Like BPOB, it also forms interactions with the phosphate groups of NAD^+. As it extends out from the binding pocket ZINC1673377 makes interactions with residues of flexible loop 1, Gly182 and Gly184. Interactions with main chain atoms of nearby residues such as Val83 and Arg84 may also be important for this ligand.

ZINC1936250

The second highest ranked hit in Virtual Screen 1, ZINC1936250, has a predicted K_i of 3.14nM. This compound was predicted to bind deeply in the L-threonine binding pocket and makes

interactions with several of the important L-threonine-binding residues. As with the previous two compounds examined, ZINC1936250, also interacts with an NAD^+ phosphate group and with Val83 and Arg84. The piperazine moiety extends outwards from the binding pocket and makes extensive interactions with the flexible loop, suggesting that this structural feature of TDH may be important in ligand binding.

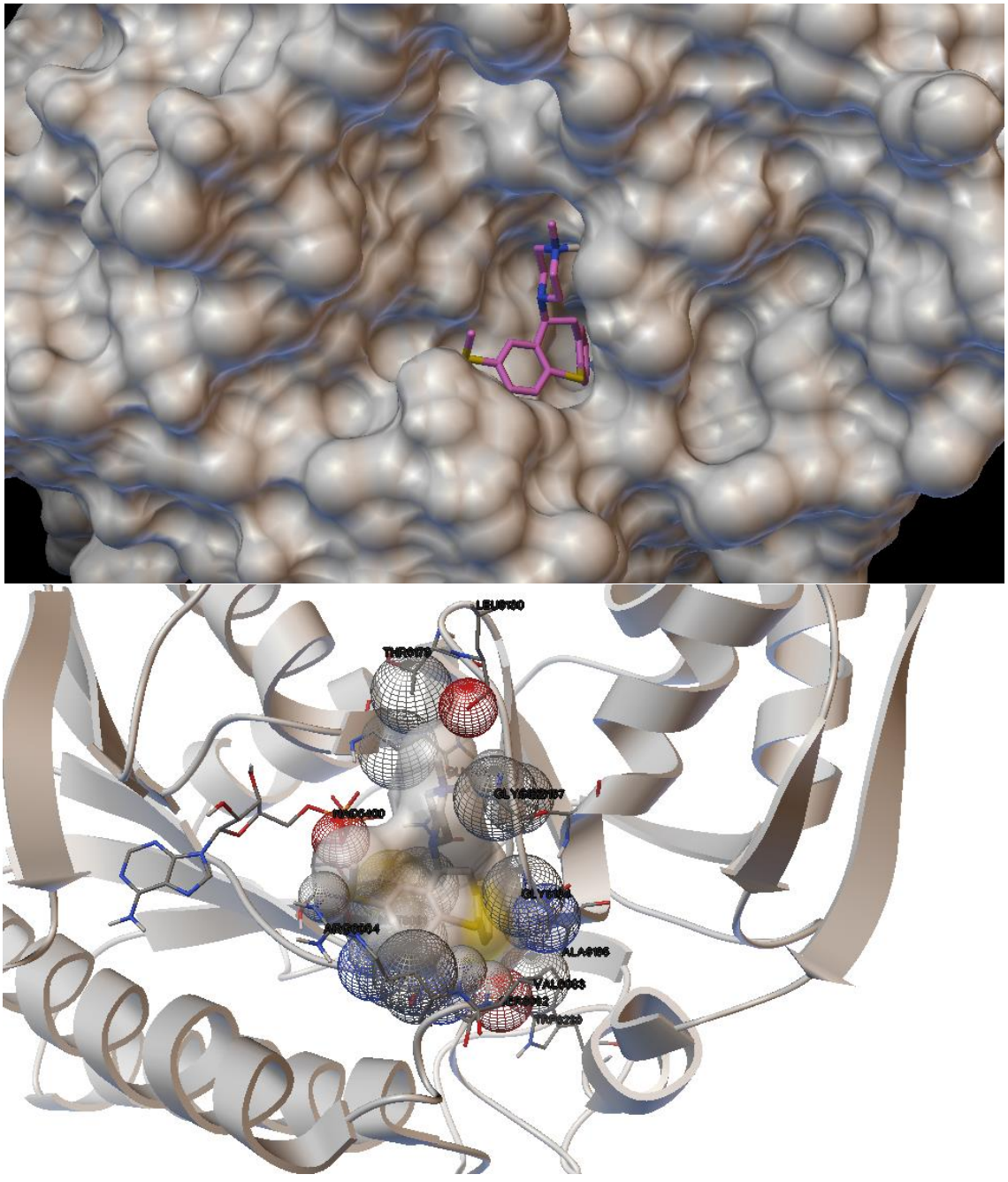


Figure 3.3.18 - predicted binding pose of ZINC1936250 depicted using ADT.

Many of the hits in Virtual Screen 1 show similar binding interactions to those described for the ligands above. Interactions with L-threonine-binding residues are important, as well as interactions with other amino acids in the binding pocket. The phosphate groups of NAD^+ interact with many of the ligands in their predicted binding poses. For molecules that bind on

the outer edge of the binding pocket, particularly larger molecules, interactions with residues of flexible loop 1 also play a role in binding.

3.3.4.2 Virtual Screen 3

Again, an examination of the binding poses of compounds with the highest predicted affinities will be used to illustrate the important interactions made with TDH by hits in Virtual Screen 3.

Alpha-solamargine

α-Solamargine was predicted to bind TDH with a binding energy of -15.03kcal/mol and a K_i of 10pM, the lowest of any compound across all four screens.

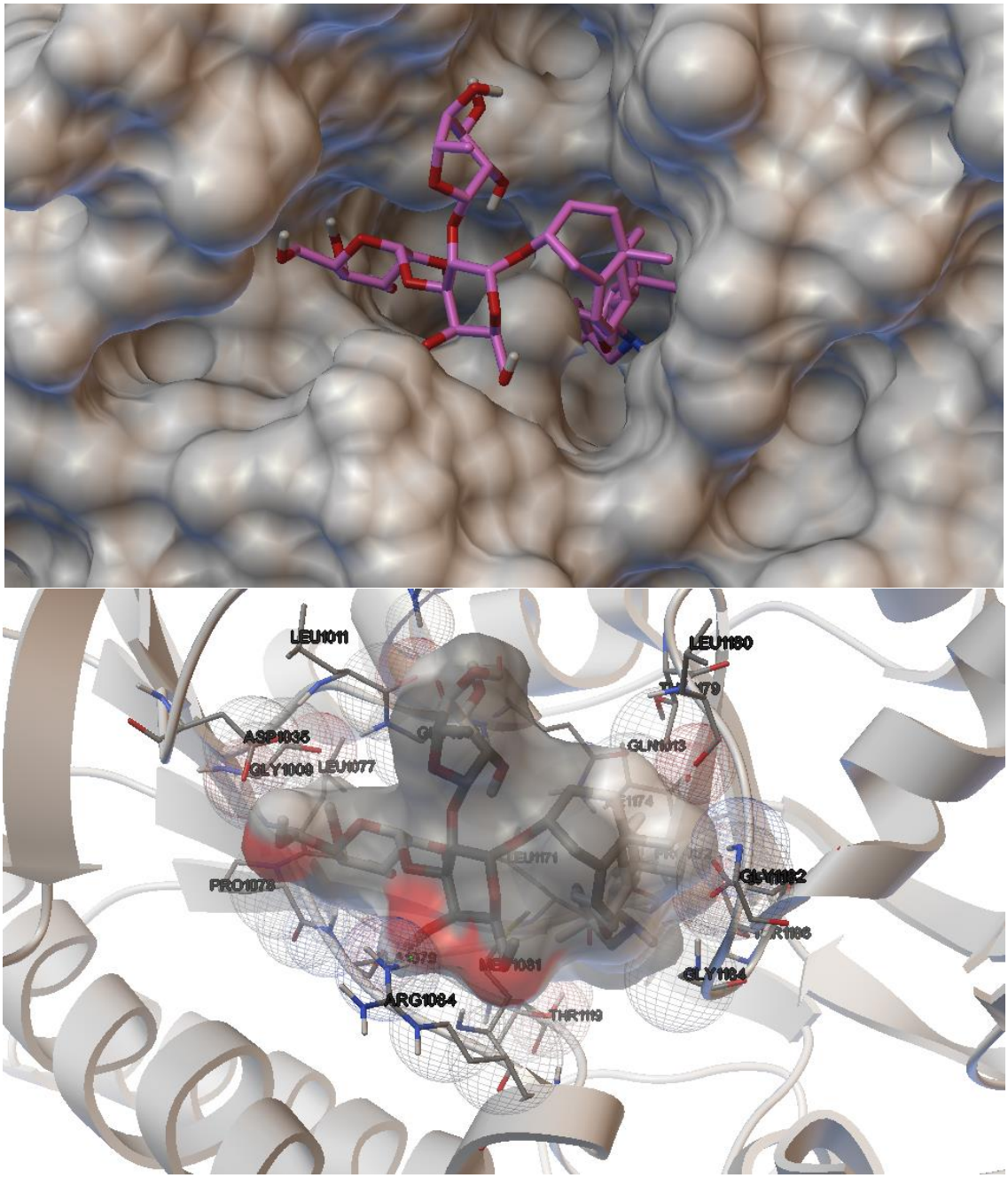


Figure 3.3.19 - predicted binding pose of α-solamargine in Virtual Screen 3, depicted in ADT.

Looking at Figure 3.3.19, it can be seen that Met81 (located near the centre of each image) acts as a boundary between the NAD^+ and L-threonine binding sites (on the left and right of the image, respectively). The large molecules that bound to both sites either bind on the inner side of Met81 in the nicotinamide binding pocket, or they bound on the outer side, as α-solamargine does. Figure 3.3.19 shows that the molecule does not bind very deeply in the L-threonine binding pocket, but it benefits from additional interactions in the NAD^+ binding site. The binding pose on the outer side of the active site means that it also makes interactions with residues of flexible loop 1, which is in the open confirmation in Virtual Screen 3.

Ancistrogriffithine A

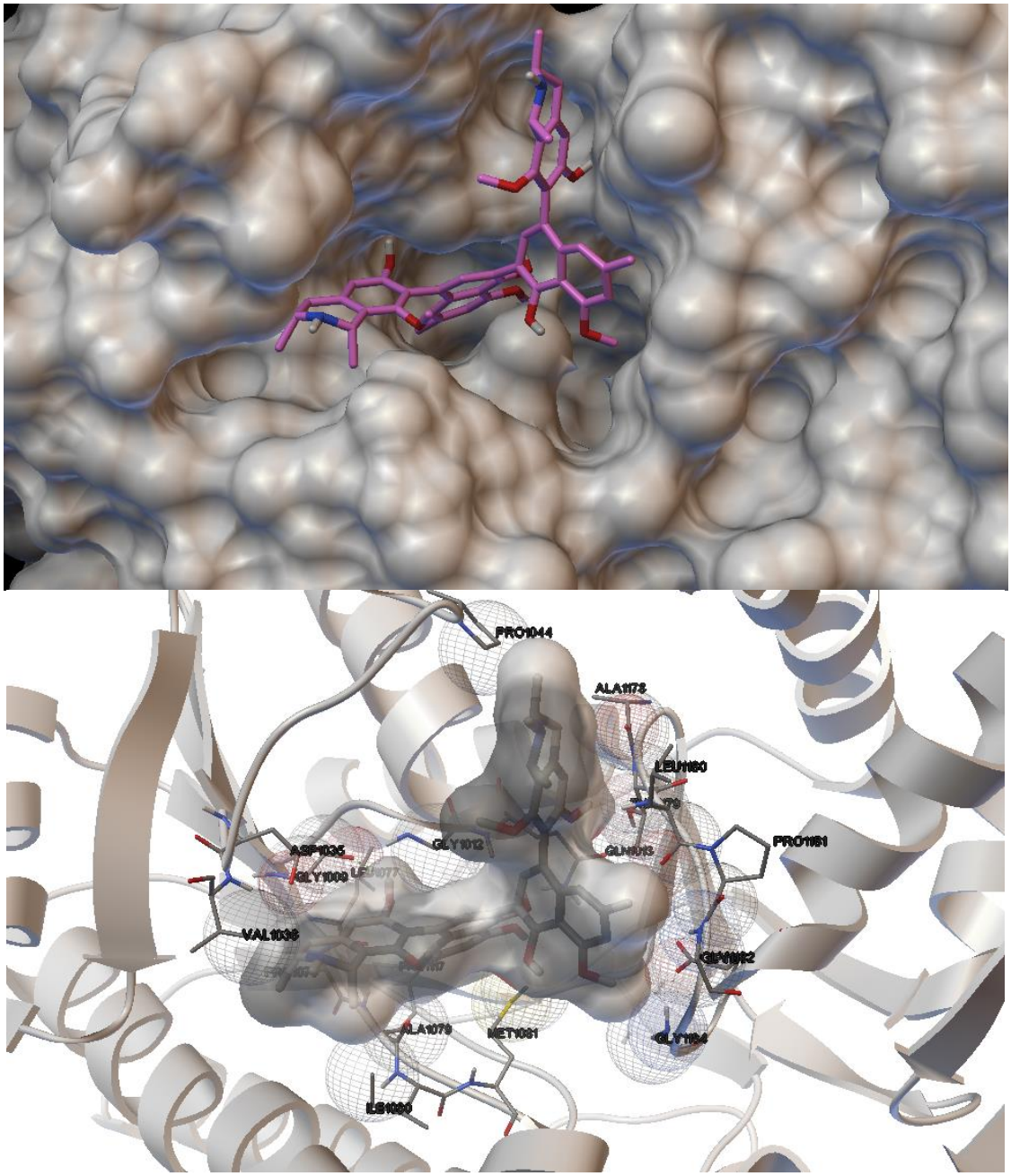


Figure 3.3.20 - binding pose of ancistrogriffithine A in AutoDock. Depicted using ADT.

Ancistrogriffithine A, another natural product of high molecular weight, achieved a predicted K_i of 100pM. In Figure 3.3.20 it can be seen to bind deeply in the nicotinamide binding pocket. Rather than interacting with the L-threonine binding site, it makes extensive interactions with flexible loop 1 and even an interaction with Pro44 of flexible loop 2.

ZINC05492794

The symmetrical molecule ZINC05492794 was predicted to bind to TDH with an equal affinity to ancistrogriffithine A. It provides a good example of a molecule binding both the NAD and L-threonine binding pockets on the inner side of Met81.

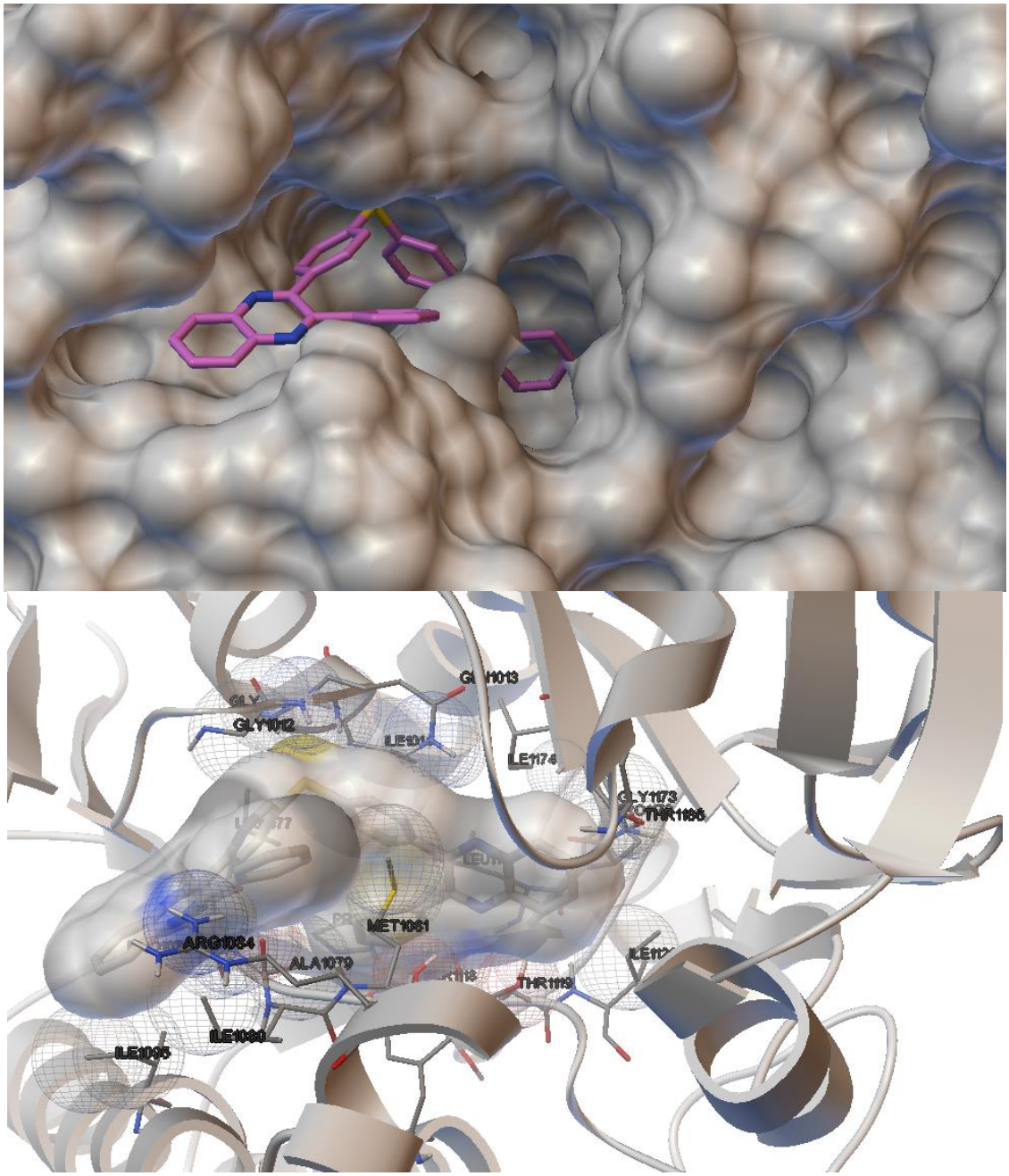


Figure 3.3.21 - predicted binding pose of ZINC05492794, depicted in ADT.

In Figure 3.3.21, ZINC05492794 can be seen binding deeply in both the L-threonine and NAD^+ binding pockets. It makes interactions with several of the important L-threonine binding residues, including Met81, Thr119, Tyr144 and Thr186. In the NAD^+ binding site, it interacts with residues that are important in binding NAD^+, such as Gly12, Gln13, Ile14 and Ile80. The binding mode of ZINC05492794 is not affected by the position of the flexible loop, which can explain the fact that this compound was also one of the top scoring hits in Virtual Screen 4.

3,23-dioxotirucalla-7,24-dien-21-al

Like ZINC05492794 and many of the other highest ranked hits from Virtual Screen 3 (see Table 3.3.5), 3,23-dioxotirucalla-7,24-dien-21-al was a hit in both Virtual Screen 3 and Virtual Screen

4. It has a predicted binding affinity of 250pM in Virtual Screen 3 and it is another example of a molecule that is predicted to bind the L-threonine and NAD binding pockets simultaneously. As it is much smaller than some of the higher ranking compounds, it binds more extensively in the L-threonine binding pocket, protruding into the nicotinamide pocket to a small degree.

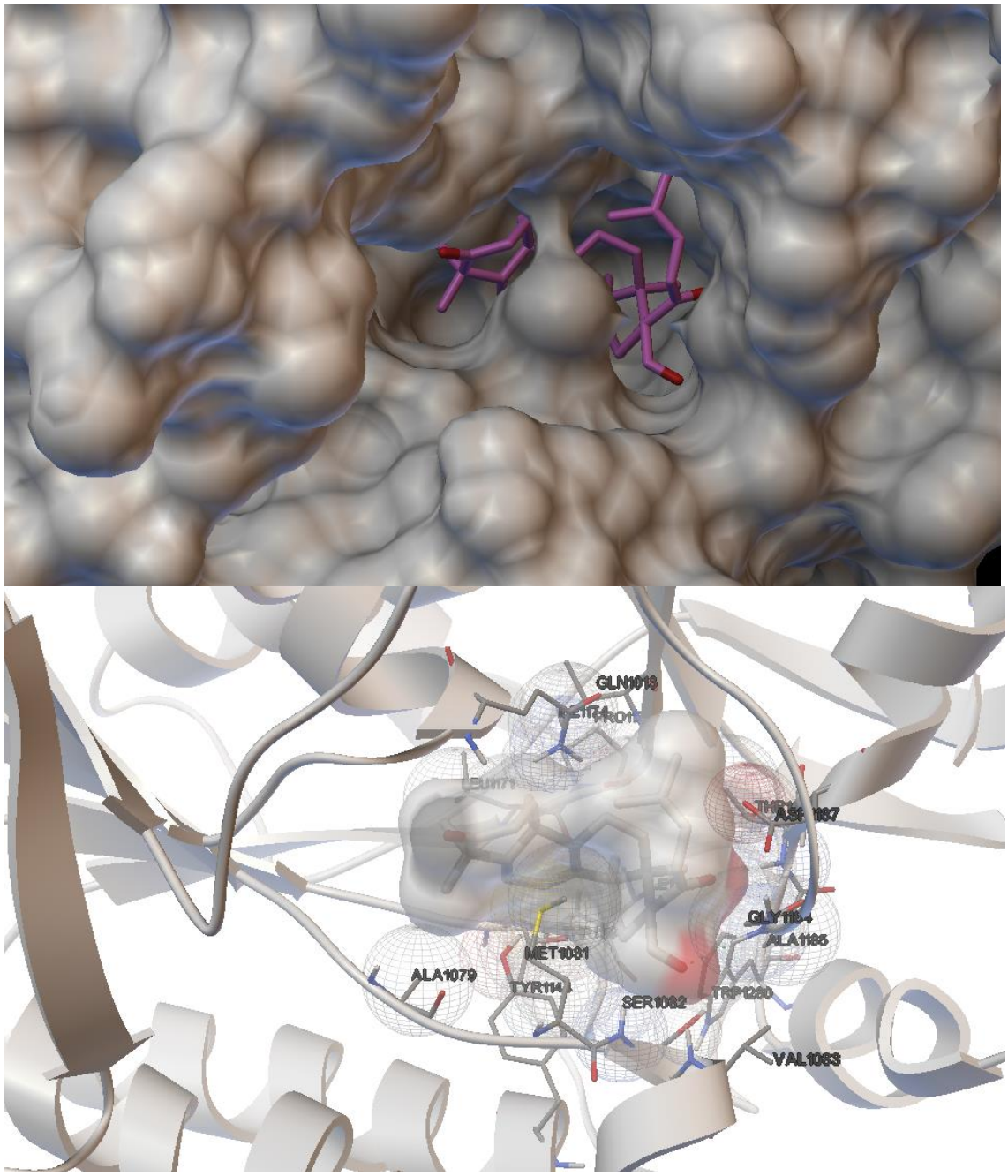


Figure 3.3.22 - predicted binding pose of 3,23-dioxotirucalla-7,24-dien-21-al in Virtual Screen 3.

Two more examples of molecules binding to both TDH substrate binding sites can be seen in Figure 3.3.23.

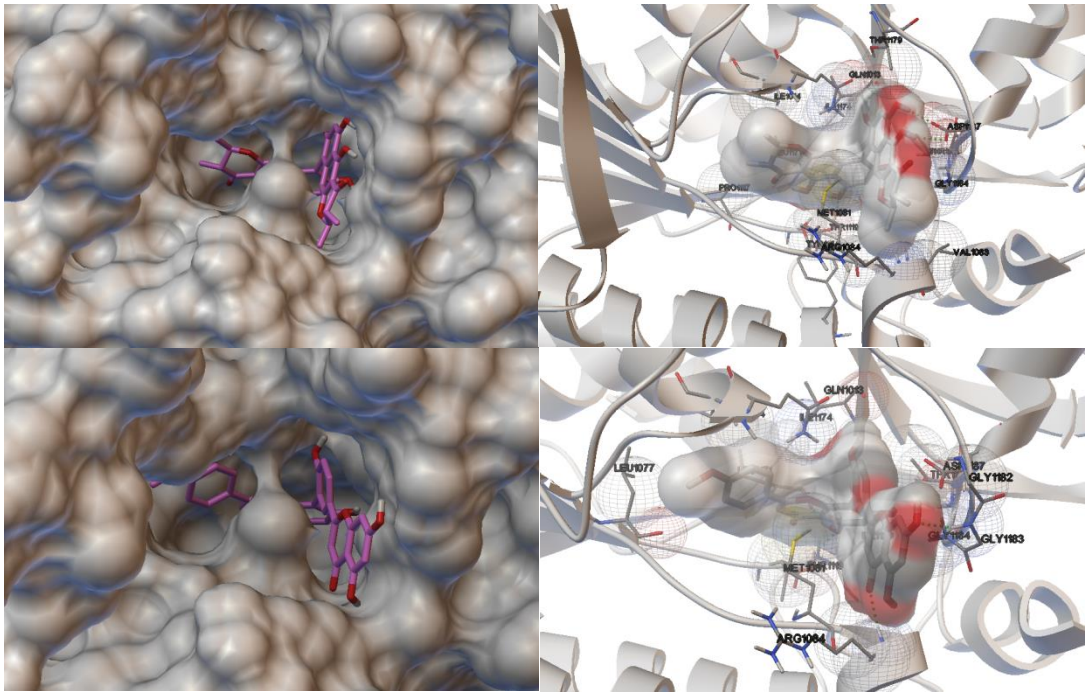


Figure 3.3.23 - predicted binding poses of ZINC17465983 (top two images) and podocarpusflavone A (bottom two images) in Virtual Screen 3. Both compounds bind deeply in the L-threonine binding pocket and nicotinamide binding pocket simultaneously.

3.3.4.3 Virtual Screen 4

In Virtual Screen 4, the high predicted binding energies of the highest ranked hits can be attributed to their binding to both substrate binding sites, as for the hits in Virtual Screen 3. In the TDH model used as the receptor in Virtual Screen 4, flexible loop 1 is in the closed conformation. For this reason, many of the hits in this screen were also predicted by the software to make interactions with residues of this loop.

Astrogaloside III

The natural product, astrogaloside III is very large and makes extensive interactions across a large area of the protein. It has a predicted K_i of 12pM.

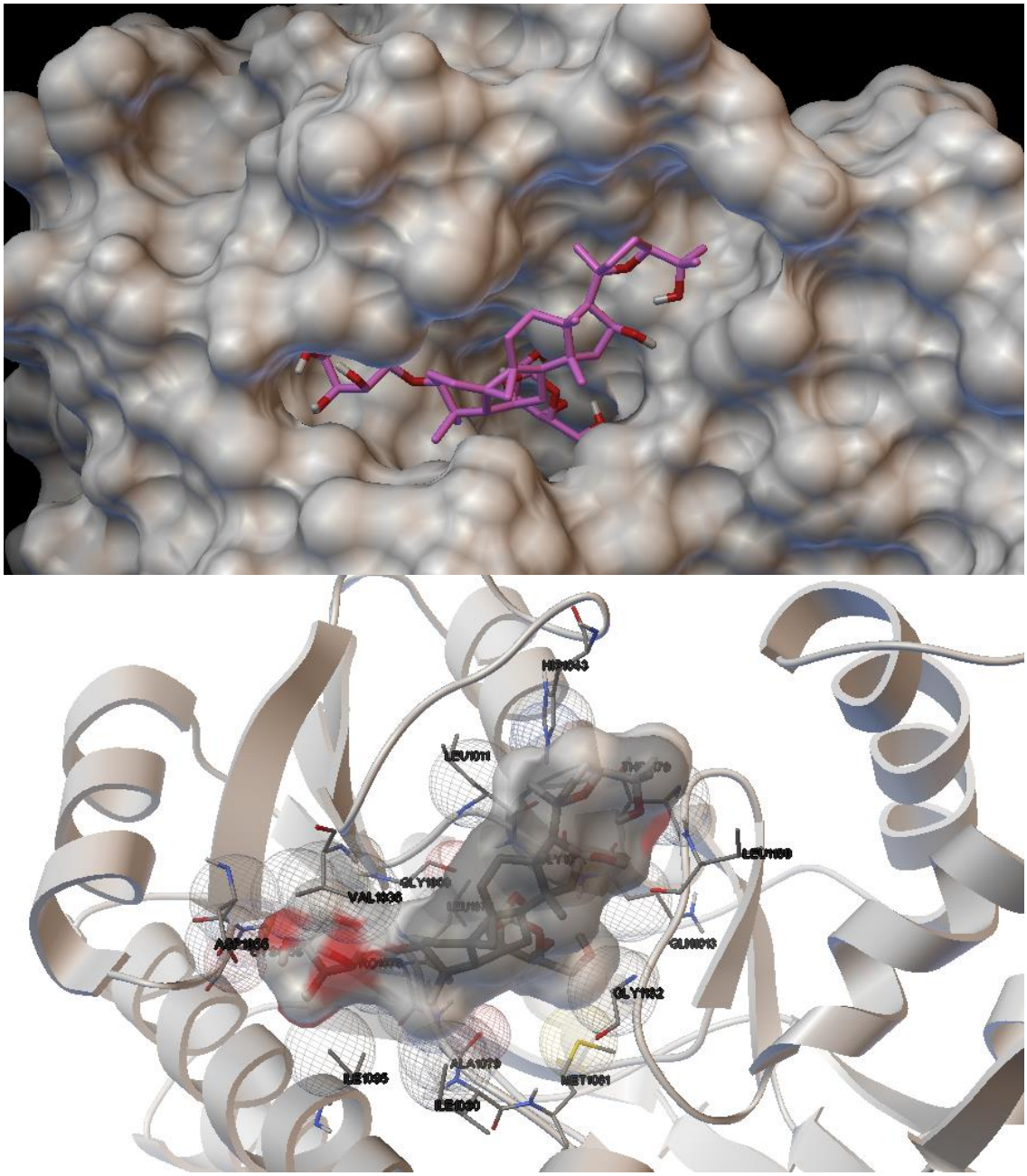


Figure 3.3.24 - predicted binding pose of astrogaloside III in Virtual Screen 4. Depicted in ADT.

The closed conformation of the flexible loop precludes binding of astrogaloside III to the L-threonine binding site. Instead, it almost completely fills the NAD^+ binding site and forms binding interactions with residues of the flexible loop from the exterior of the protein, rather than from within the L-threonine binding site.

Stigmasterol D-glucoside

Stigmasterol D-glucoside (predicted K_i = 210pM) offers an example of a compound predicted to bind in both substrate sites simultaneously in this screen.

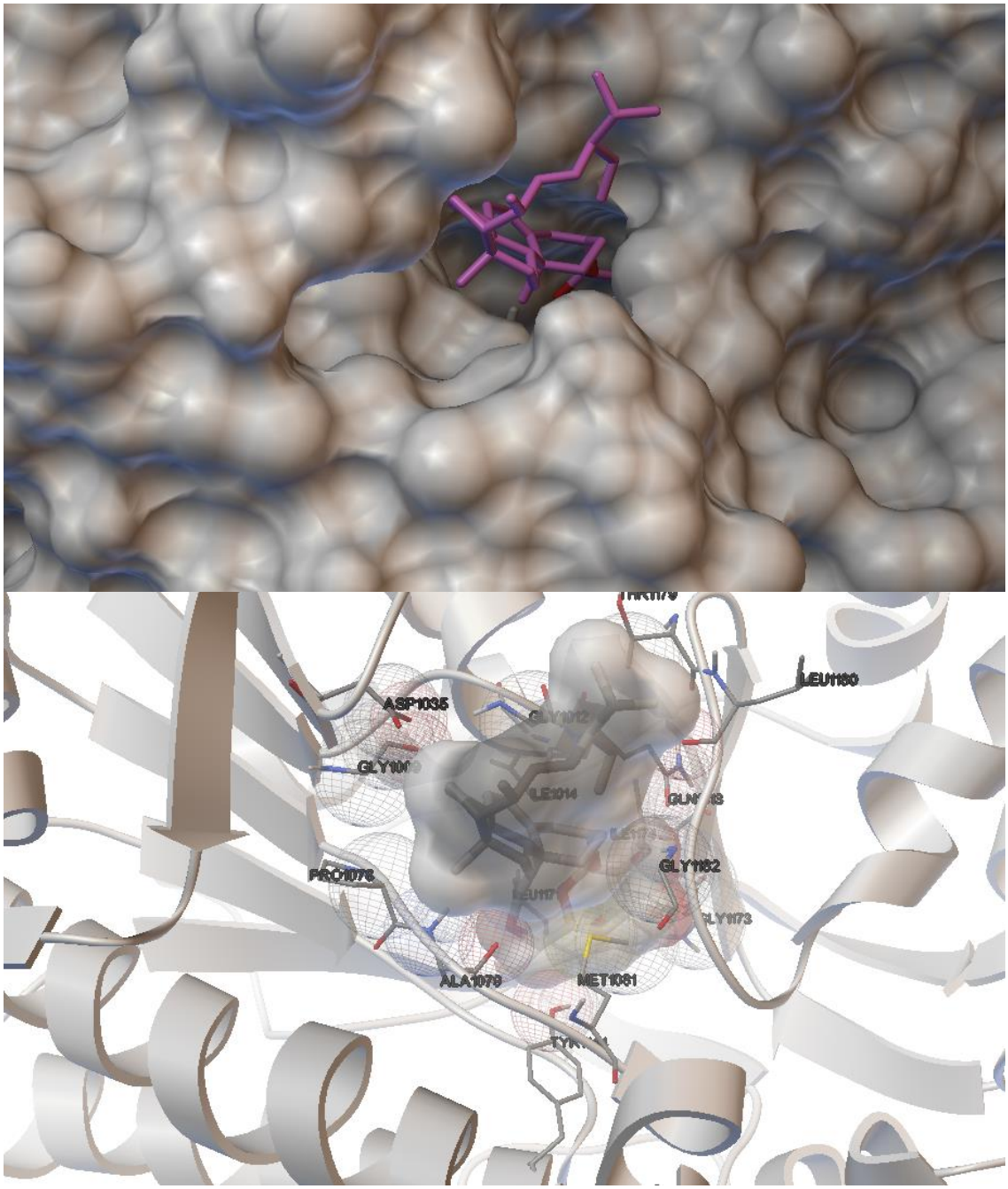


Figure 3.3.25 - predicted binding pose of stigmasterol D-glucoside in Virtual Screen 4. Depicted in ADT.

The steroid moiety of stigmasterol D-glucoside can be seen to bind in the nicotinamide pocket, behind Met81. The sugar moiety then protrudes into the L-threonine binding pocket and interacts with the threonine-binding residue, Tyr144. Here, it also makes many interactions with the closed loop. The hydrophobic tail at the other end of the molecule interacts with flexible loop residue Thr179 in a similar manner to astrogaloside III.

By observing these predicted intermolecular interactions it can be seen that the availability of both the NAD^+ and L-threonine binding sites determines which molecules bind TDH, how they bind TDH, and the combined strength of all hydrogen bonds and Van der Waals interactions. The top scoring compounds in all screens interact with the same residues that are important for binding the substrates. The open loop, in its different conformations, affects binding to the threonine-binding site, whilst also providing more bonding partners itself.

3.4 *In vitro* Screening

3.4.1 Compound libraries

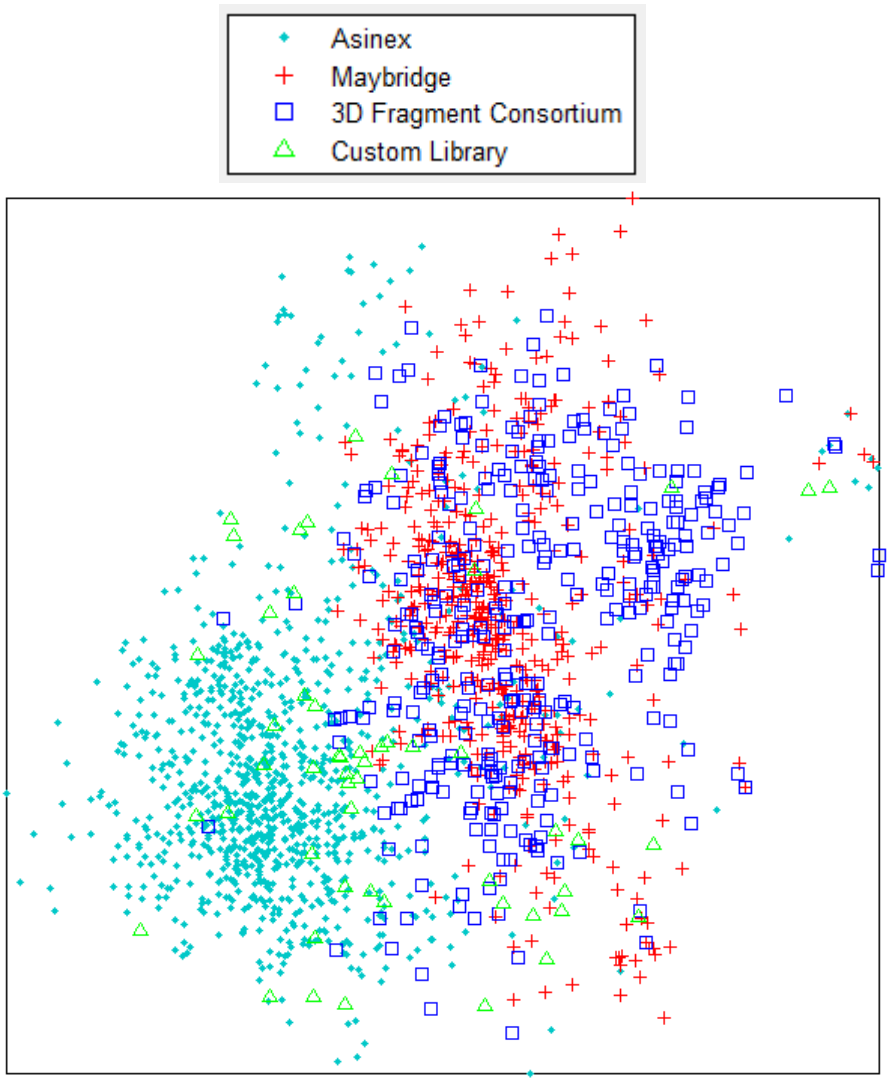


Figure 3.4.1 - visualisation of the diversity of all compounds in the *in vitro* screening library. Each point represents a single compound in the screening library, and a different shape and colour is used depending on which subset of the library the compound belongs to. The relative positions of the points are dependent on the differences between their chemical structures.

The visualisation of the compounds in the *in vitro* screening library (Figure 3.4.1) suggests that a reasonable degree of diversity was achieved. The fragments appear to be concentrated around the centre of the diagram. This is probably due in part to their smaller size. The compounds are well distributed vertically, suggesting that a good degree of diversity exists within the limits of compounds that size. The points representing the compounds from the 3D fragment consortium are noticeably more dispersed than those of the Maybridge library. This is probably a reflection of the specific design of that library, which aimed to develop more compounds with 3D geometries, as opposed to the flat structures often found in commercial fragment

libraries(243). Figure 3.4.1 shows that the compounds in the custom library occupy diverse regions of chemical space and there are only a few groupings of similar molecules. Conversely, the compounds of the Asinex library appear to be highly concentrated in one corner of the diagram, suggesting that these molecules do not cover a diverse range of chemical features. There is the possibility that their concentration in this part of the diagram is a consequence of those molecules having a molecular weight that is very different from the fragments represented in other areas of the diagram. However, this pattern also raised the concern that the virtual screen carried out by UCL ChemiBank using the program GOLD, identified hits with a limited range of molecular structures. This would have been in contrast with the results of the AutoDock virtual screens, which found hits with diverse chemical structures.

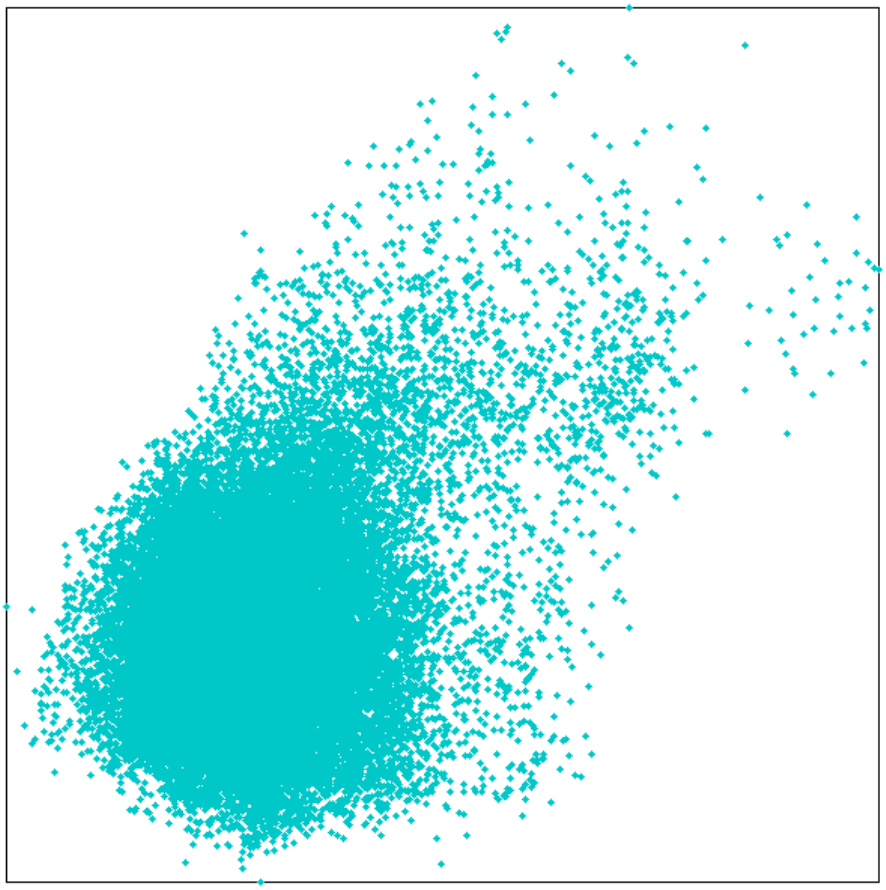


Figure 3.4.2 - visualisation of the chemical diversity of all compounds in the Asinex library (20,000 compounds) screened by UCL ChemiBank, using GOLD.

Figure 3.4.2 shows a data visualisation of the original 20,000 compounds in UCL ChemiBank's Asinex library. The compounds are again concentrated in the same area of the diagram, so the result in Figure 3.4.1 is not a consequence of a bias introduced by Virtual Screening. Rather, this confirms that GOLD found hits with diverse features from the original pool of compounds. This is similar to the performance of AutoDock in the virtual screens described above.

Table 3.4.1 - mean chemical terms describing the features of the compounds in the *in vitro* screening libraries.

	Mean Chemical Term Value								
Library	**MW**	**LogP**	**TPSA**	**Hydrogen-bond Acceptors**	**Hydrogen-bond Donors**	**Number of Rotatable Bonds**	**Lipinski Rule of Five Compliant?**	**Less than 1 Rule of Five Violation?**	**Congreve Rule of Three Compliant?**
Asinex	468.1	3.85	93.51	5.2	1.3	7.4	50.7% (507 of 1000)	87.7% (877 of 1000)	1.5% (15 of 1000)
Maybridge	189.1	1.31	45.12	2.6	1.0	1.8	100.0% (500 of 500)	100% (500 of 500)	75.4% (377 of 500)
3D Fragment Consortium	195.1	0.21	47.20	2.5	0.9	1.9	100% (358 of 358)	100% (358 of 358)	72.1% (258 of 358)
Custom Library	381.5	3.09	89.95	5.0	2.5	3.4	61.1% (33 of 54)	90.7% (49 of 54)	3.7% (2 of 54)
All	341.6	2.48	72.08	4.0	1.2	4.8	73.1% (1398 of 1912)	93.3% (1784 of 1912)	34.1% (652 of 1912)

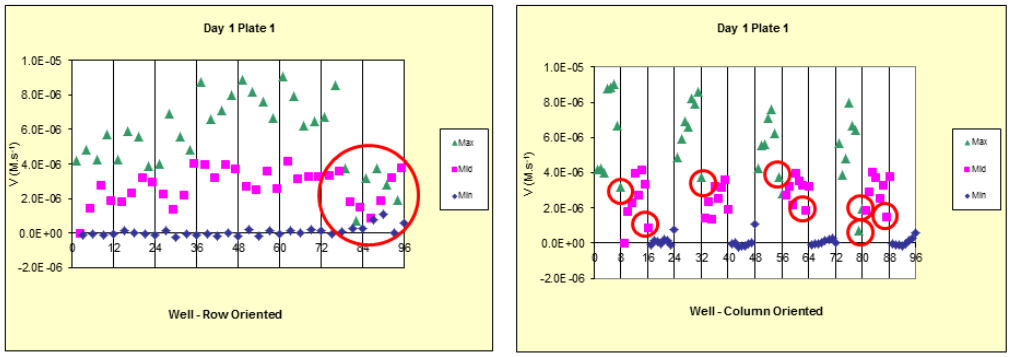
The Asinex compounds had the largest molecular weight on average, of the compounds screened in the *in vitro* inhibition assay. Despite being designed for "lead-likeness"(244), approximately 50% of the compounds in the Asinex library violate Lipinski's rule of five. However, the majority of compounds have one violation of the rule or less. The majority of the compounds in the custom library also had one Rule of Five violation or less.

The two fragment libraries share very similar characteristics in regards to their molecular weights, hydrogen bonding partners and rotatable bonds. However, there is a significant difference between their LogP values. The higher LogP of the Maybridge compounds reflects the greater presence of aromatic ring structures in the Maybridge library, compared to the 3D fragment consortium library. As is to be expected, 100% of the fragments obey the Lipinski Rule of Five. The majority of these compounds also obeyed the Congreve rule of three for fragment screening libraries.

3.4.2 Assay Validation

The Plate Uniformity Test (PUT) developed by NIH and Eli Lilly researchers(203) was a useful diagnostic test for indicating the reliability of the *in vitro* screening assay (see Section 2.4.2 for a description of the PUT assay). The visual information and the related calculated statistics (available in a Microsoft Excel file provided by the authors of the test) not only helped to identify problems with the assay, but also advised on the appropriateness of the assay design. The signal window and the Z-score are two related parameters that describe the separation between the maximum and minimum signals in an assay. Calculation of these parameters helped to judge the ability of the assay to distinguish inhibitors from false positives. Figure 3.4.3 shows examples of the graphs and statistics calculated by the PUT Microsoft Excel file.

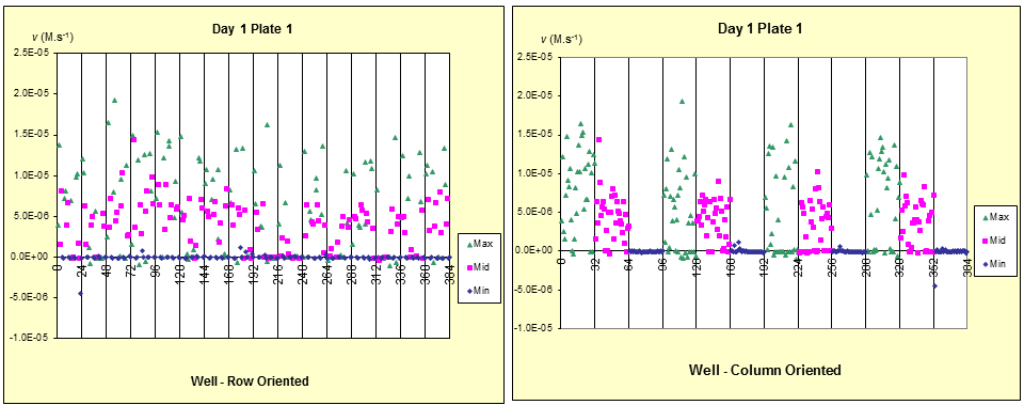
A



Circled points (O) relate to where enzyme lost activity towards the end of the assay.

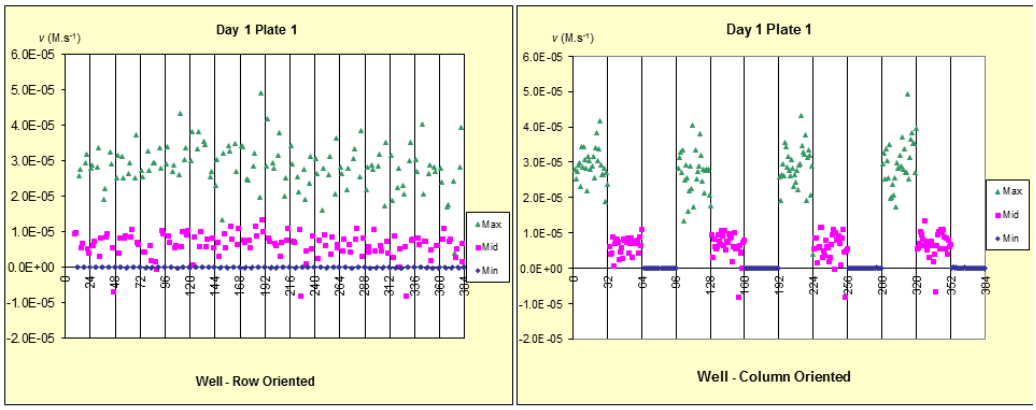
Signal Window = 0.33
Z'-score = 0.09

B



Signal Window = 0.33
Z'-score = 0.09

C



Signal Window = 1.57
Z'-score = 0.34

Figure 3.4.3 - graphical interpretation of the different signals recorded in the PUT. The "Max" signal, relating to un-inhibited TDH activity, is represented by green points. The "Mid" signal, relating to 30-70% activity/inhibition, is represented by pink points. The "Min" signal, which relates to zero enzyme activity is represented by dark blue points. The graphs under 'A' show an assay where the enzyme lost activity towards the end of the assay. Under 'B', the results of an assay where the Max and Mid signals were highly variable are shown. Under 'C', the results of a PUT using the final assay format are shown. Under each letter, graphs arranged according to microplate rows and columns are displayed to help identify any effects related to wells on the edge of the plate.

Figure 3.4.3 shows the results of the PUT of the final assay format, which used a reaction buffer with 1mM KCl included. This test was close to the published limits for a screening assay (Signal Window $\geq$ 2.0 or Z'-score $\geq$ 0.4) where one concentration of each inhibitor is tested per plate. The importance of this during hit identification and hit confirmation will be discussed below.

3.4.3 *In vitro* screening hits

In vitro screening, was carried out as described in Section 2.4.2. Screening with the Maybridge library identified 44 compounds that met the hit criteria of exhibiting at least 50% inhibition of TDH in two separate assays. This represents a hit rate of 8.8%. After hit confirmation studies, only seven compounds were confirmed as TDH inhibitors. The confirmed hit rate was therefore only 1.4% and the false positive rate was approximately 7.4%.

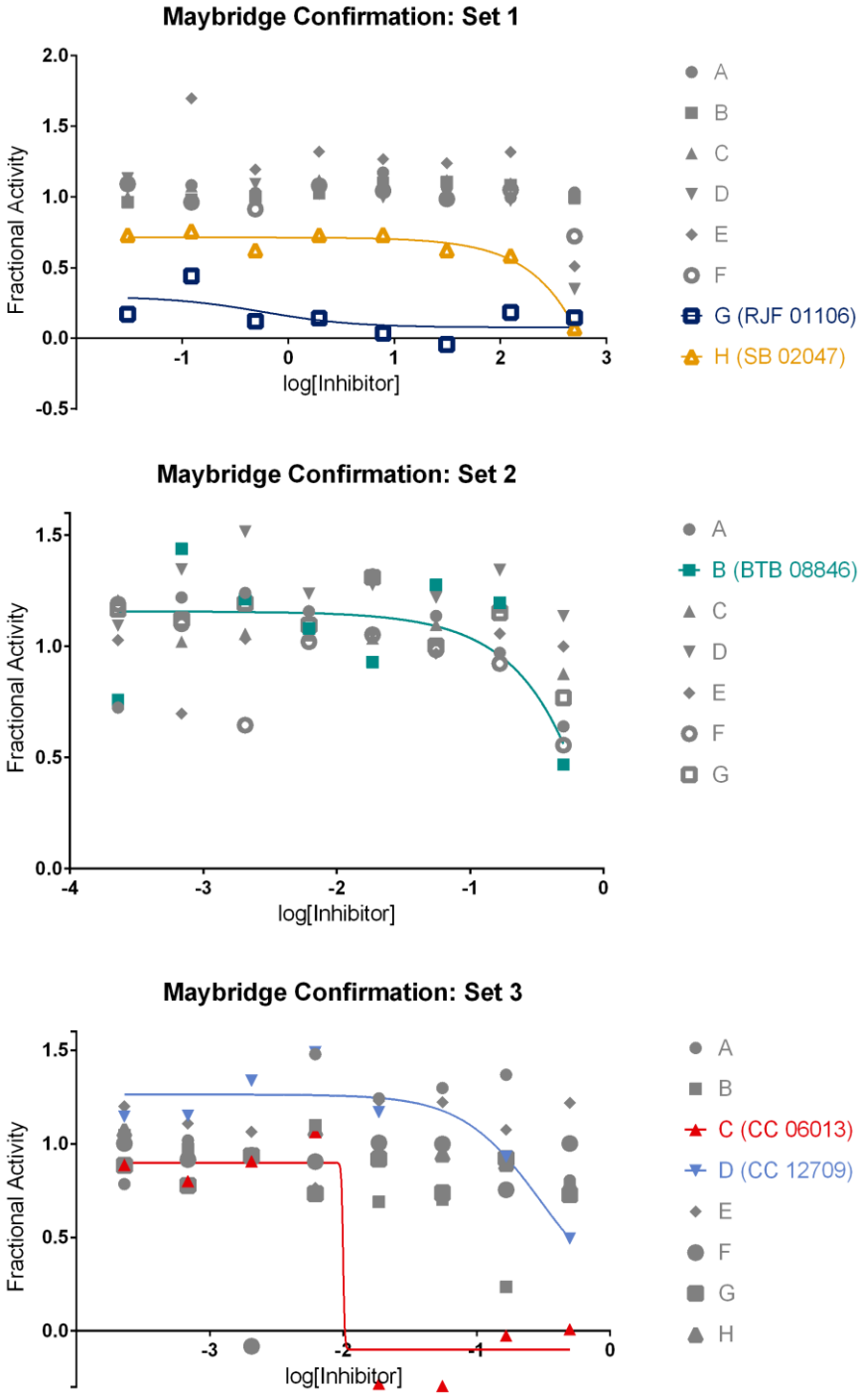


Figure 3.4.4 - plots of fractional TDH activity against the log of inhibitor concentration tested in hit confirmation assays for Maybridge compounds. Hits from the primary screen were tested across a range of concentrations (three- or four-fold dilutions). Points relating to confirmed hits are coloured and dose-response curves are fitted to the data. Points relating to false positives are coloured grey.

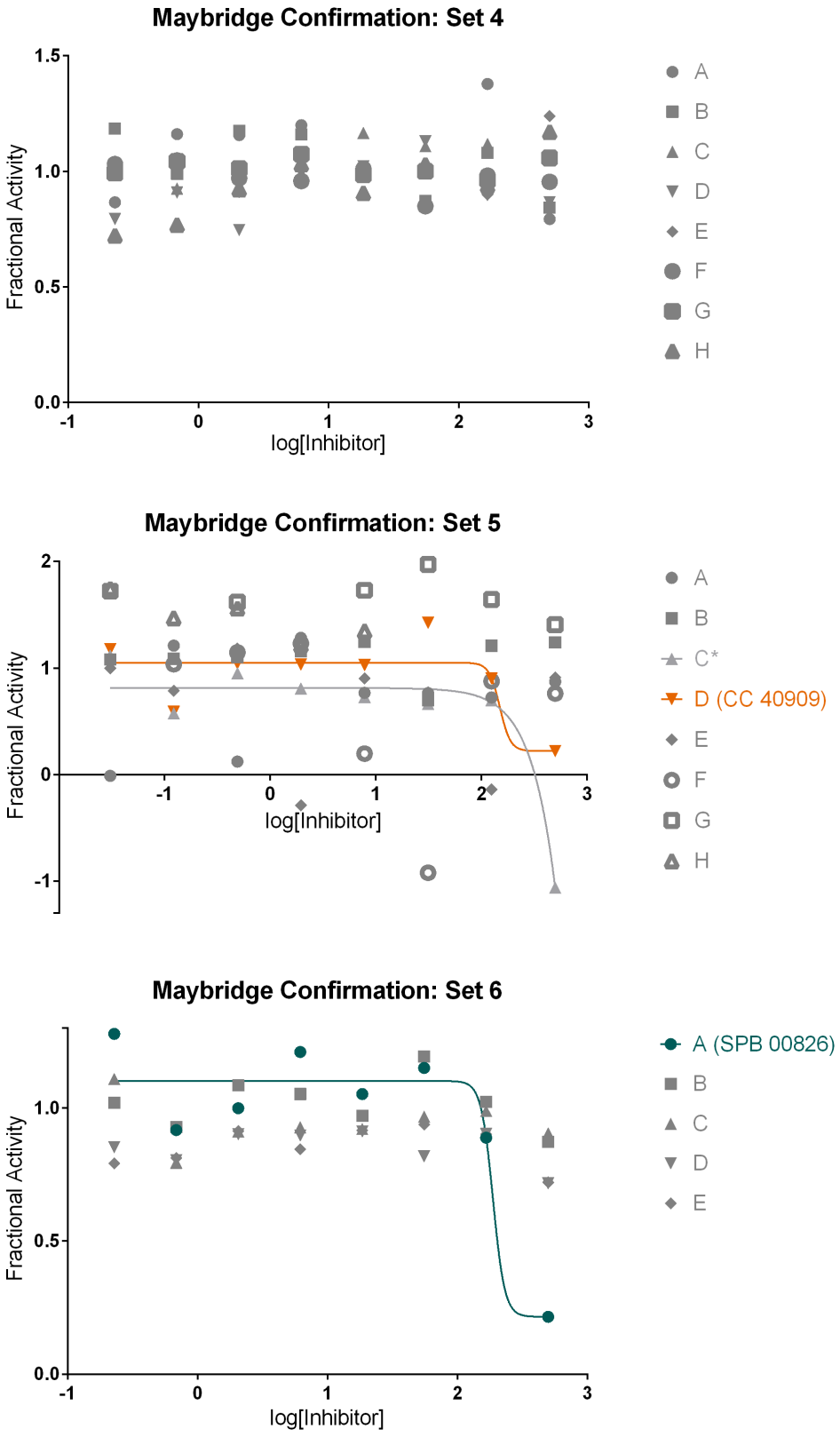


Figure 3.4.5 - plots of fractional TDH activity against the log of inhibitor concentration tested in hit confirmation assays for Maybridge compounds. Hits from the primary screen were tested across a range of concentrations (three- or four-fold dilutions). Points relating to confirmed hits are coloured and dose-response curves are fitted to the data. Points relating to false positives are coloured grey.

Figure 3.4.4 and Figure 3.4.5 show the confirmation of hits as those showing a concentration-dependent inhibition, starting at least 50% at 500µM. Of the seven confirmed inhibitors, three compounds, RJF 01106, SB 02047 and CC 06013, were significantly more potent.

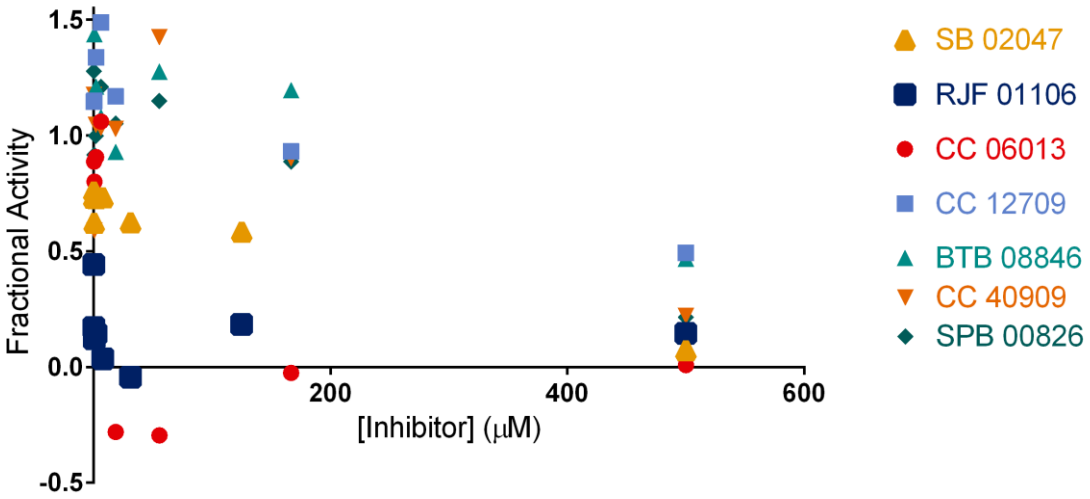


Figure 3.4.6 - plot of fractional activity as a function of the concentration of confirmed hits from the Maybridge library.

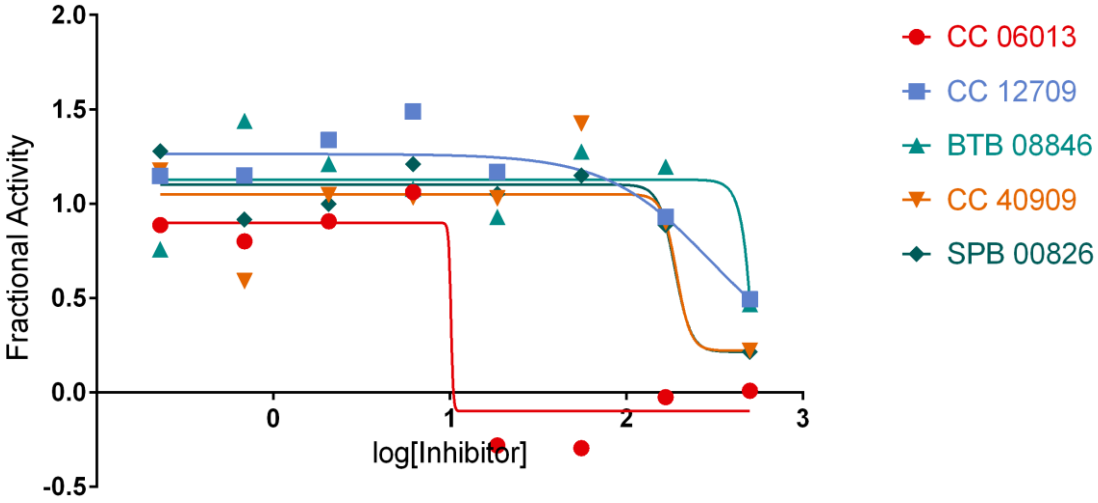


Figure 3.4.7 - plot of fractional activity as a function of the log concentration of five confirmed hits from the Maybridge library. Dose-response curves are fit to the data.

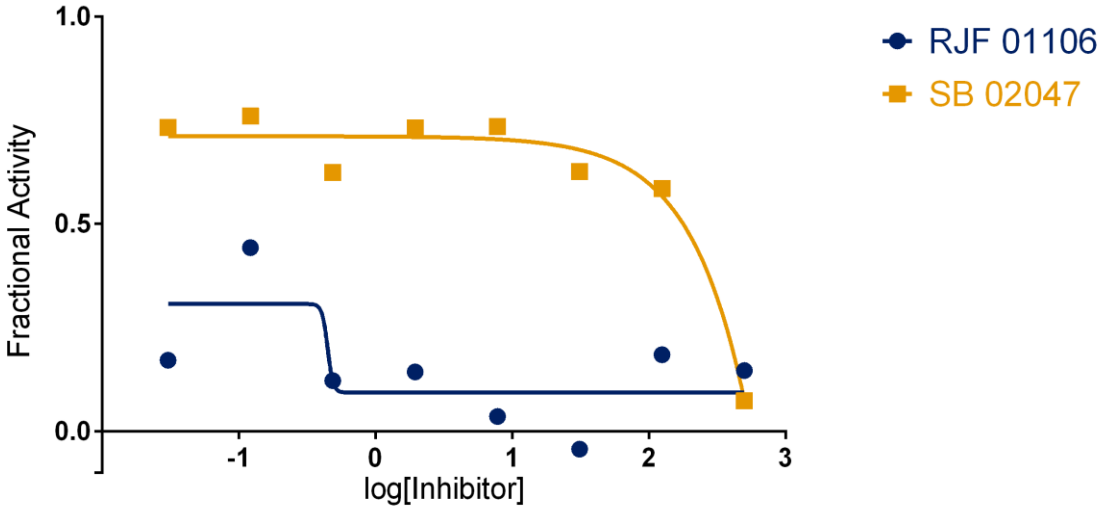


Figure 3.4.8 - plot of fractional activity as a function of the log concentration of two potent confirmed hits from the Maybridge library. Dose-response curves are fit to the data.

It can be seen that several of the fractional activity values in the figures above exceed 1.0. This is a consequence of the suboptimal assay used during these experiments. The mean enzyme activity decreased gradually with time, so that activity values were lower in both control and experimental wells towards the end of the assay. As the data have been normalised to the same scale based on the mean activity of the controls in each set of hit confirmation assays, several fractional activity values where there was no enzyme inhibition are above 1.0 and some are below one. With knowledge of this potential source of error, the observation of concentration-dependent inhibition was a useful criterion to distinguish true hits from false positives.

Figure 3.4.9 shows the results of confirmation assays for hits from the 3D fragment consortium library.

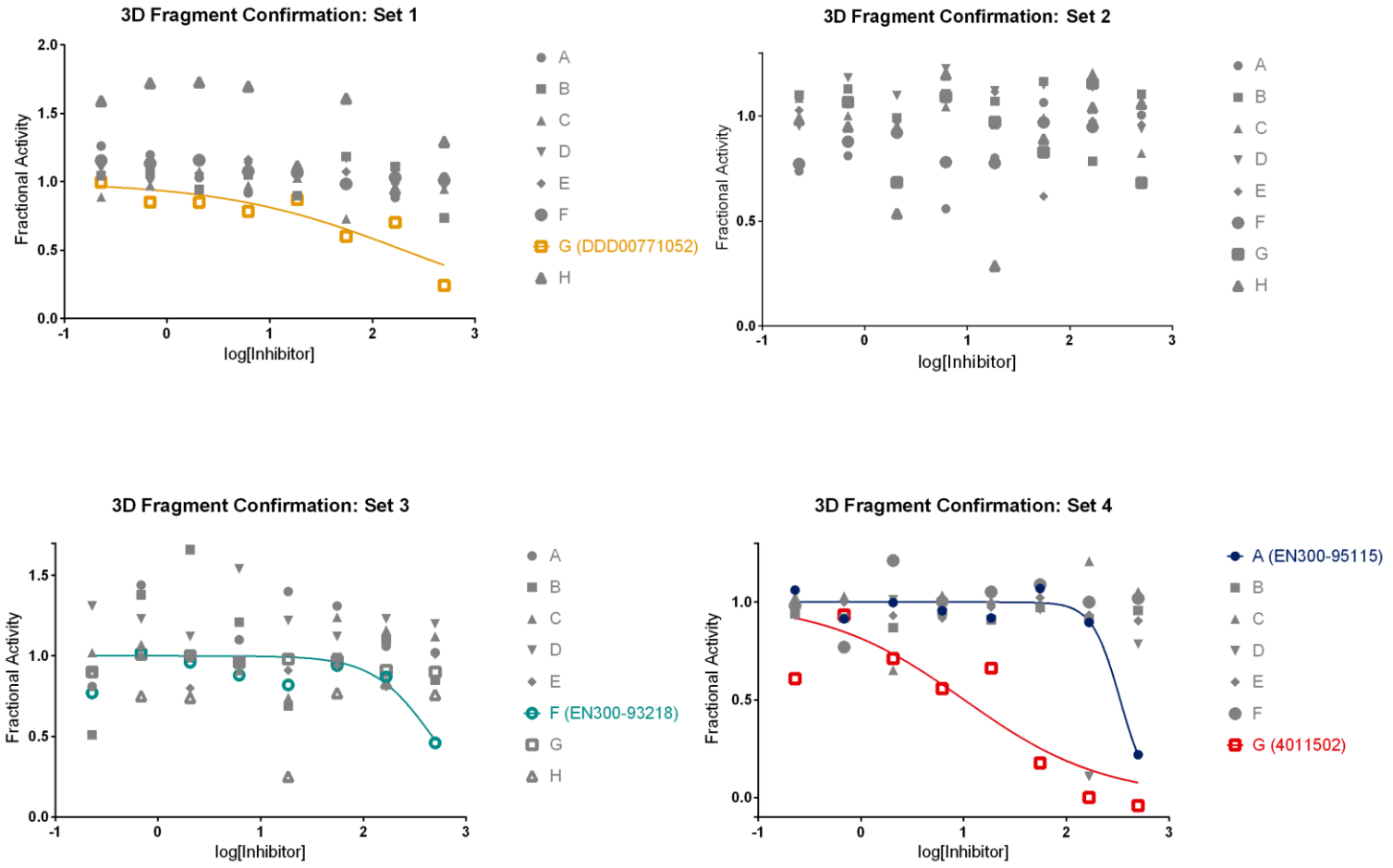


Figure 3.4.9 - plots of fractional activity against the log concentration of hits from the 3D Fragment Consortium Library. Four compounds were confirmed as inhibitors and dose-response curves were fit to the data points representing those compounds.

Of 31 hits identified from the primary assay, only four were confirmed as hits. This signifies a hit rate of 1.1%, with a false positive rate of 7.5%.

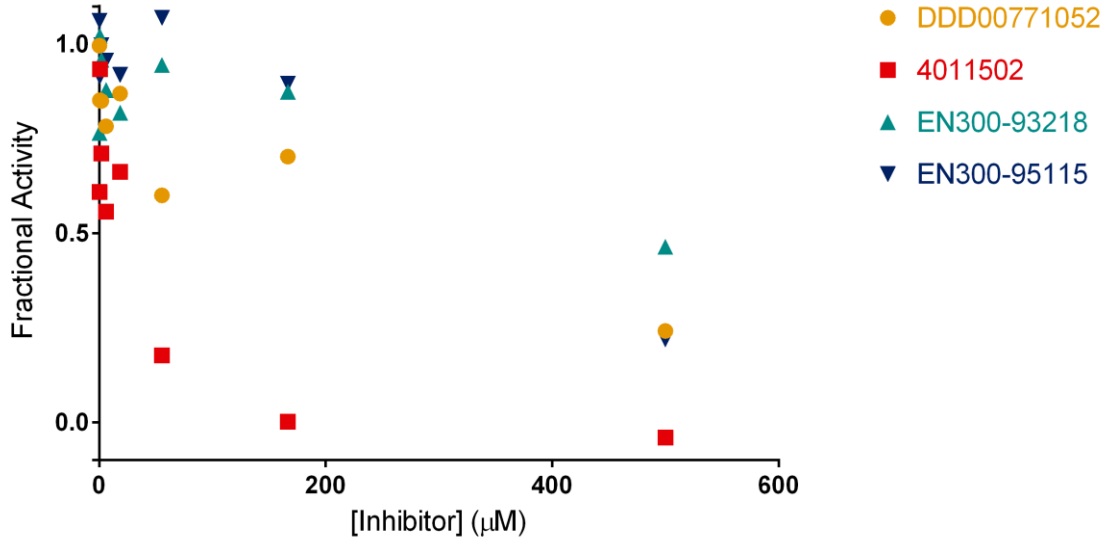


Figure 3.4.10 - plot of fractional activity as a function of the concentration of confirmed hits from the 3D Fragment Consortium library.

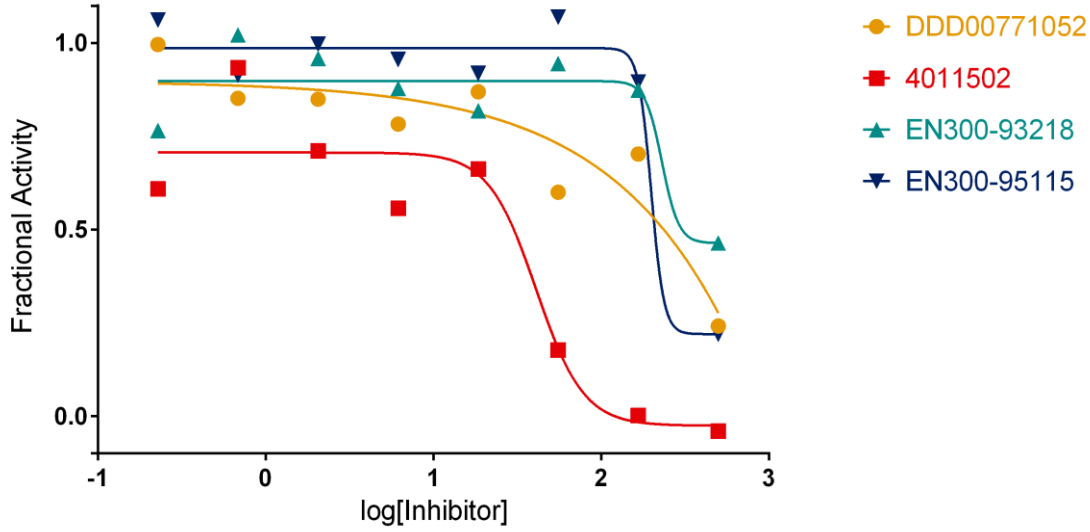


Figure 3.4.11 - plot of fractional activity as a function of the log concentration of confirmed hits from the 3D Fragment Consortium library.

As with the Maybridge fragments, two of the hits, DDD00771052 and 4011502, appear to be significantly more potent than the other inhibitors.

Custom Library

Of the eight compounds originally identified as hits from the Custom library, seven met the hit confirmation criteria of concentration-dependent inhibition of at least 50% at 500µM, and significant inhibition at 100µM. In contrast to the previous assays, the false positive rate was very low.

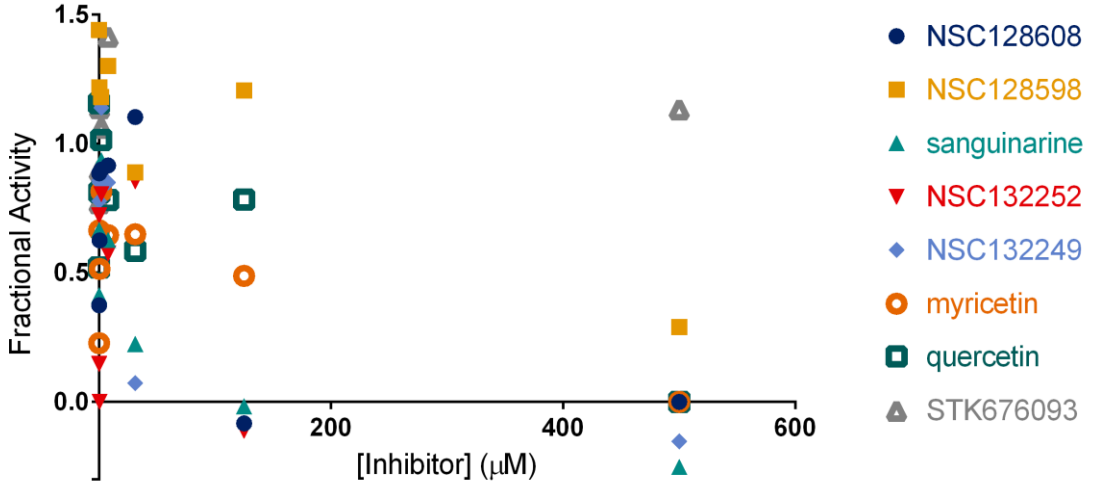


Figure 3.4.12 - plot of fractional activity against the concentration of custom library compounds in the hit confirmation assay. The only false positive, STK676093 is coloured grey.

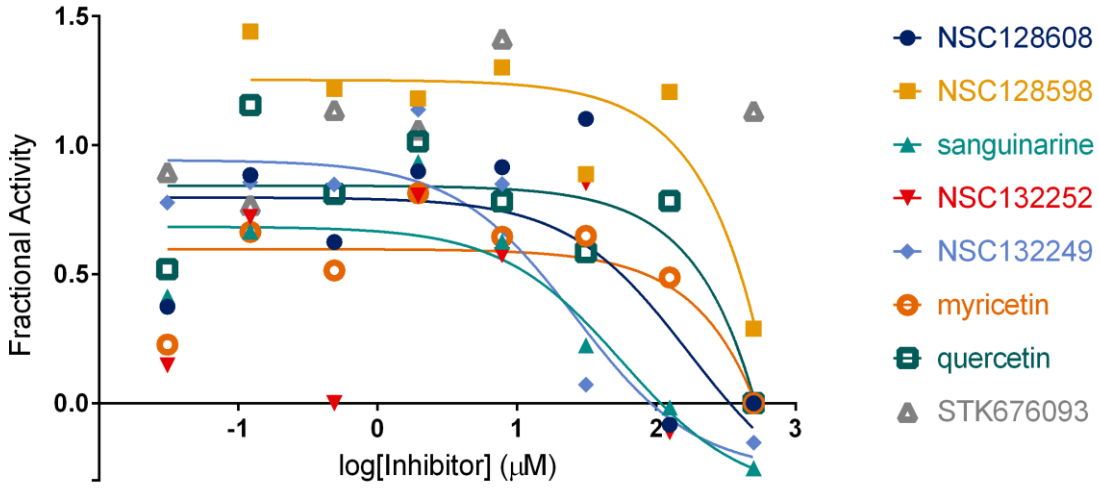


Figure 3.4.13 - plot of fractional activity as a function of the log of the concentration of hits from the Custom library. Dose-response curves are fit to the data representing the confirmed hits.

Two compounds of interest were tested for inhibition in separate assays: BPOB and Qc1. The L-threonine analogue, BPOB, which was identified through virtual screening (see above), was tested separately as prior assays had confirmed its *in vitro* activity under different conditions. BPOB was classified as a hit, as it was able to show significant inhibition at 500µM in the primary screening assay conditions. The inhibitor of *Mm*TDH, Qc1, failed to show inhibition in

the primary assay, but showed significant inhibition (6-60%) at higher concentrations of L-threonine and NAD^+. Both of these inhibitors were included as hits, but with knowledge gained from outside of the screening assay. To consider the hit rate in regards to the ability of the screening assay to identify hits, only seven compounds were identified initially, amounting to a hit rate of 13.0%. If we consider the compounds discovered through separate assays, the hit rate is 16.7%, as reported in Figure 3.4.17.

Asinex

Of fourteen hits from the primary screen of the Asinex compounds, seven were confirmed as inhibitors with concentration-dependent inhibition of at least 50% at 100μM.

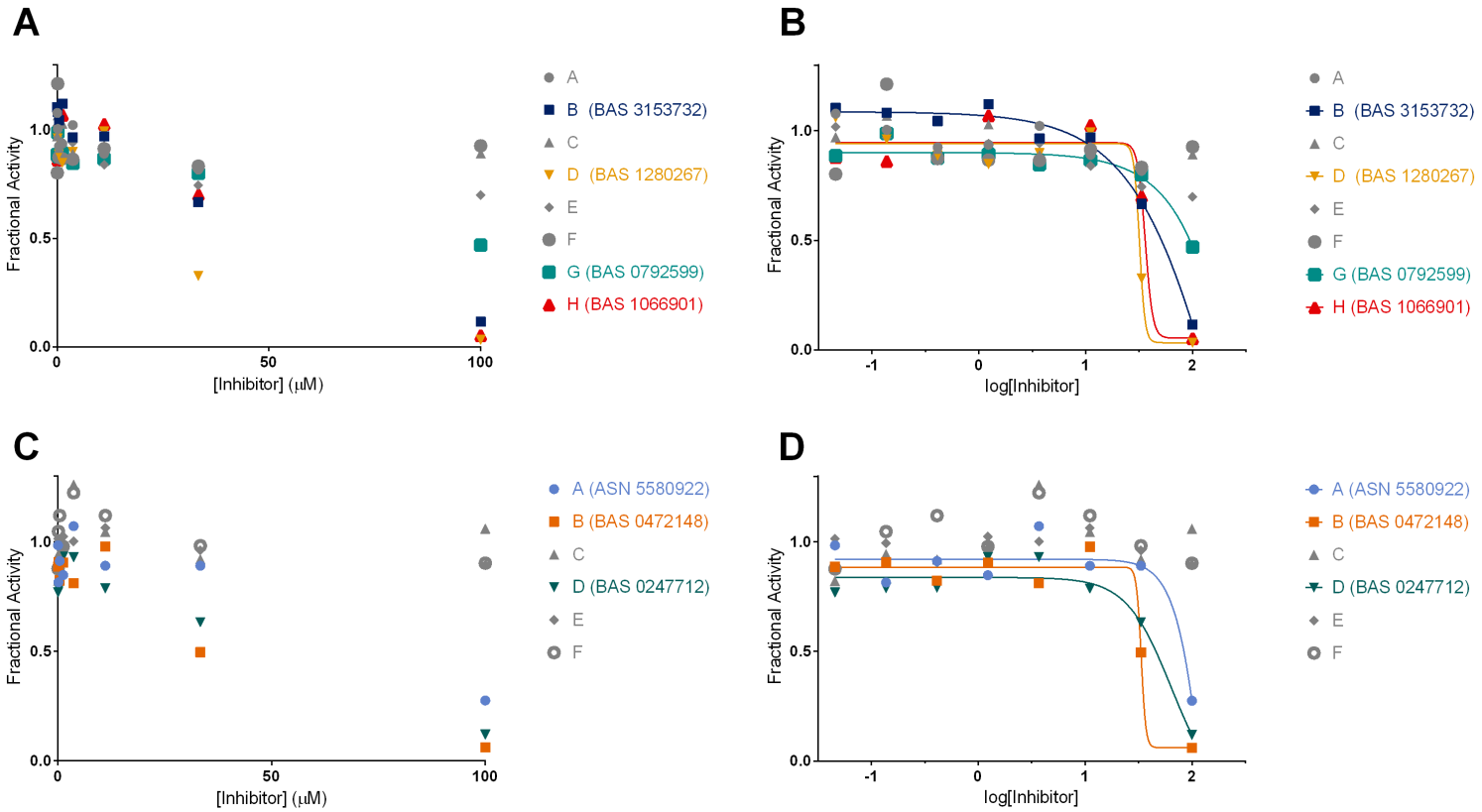


Figure 3.4.14 - plots of fractional activity as a function of three-fold dilutions of hits from the Asinex library. Plots are shown as a function of inhibitor concentration (left, A and C) and log inhibitor concentration (right, B and D), with associated dose-response curves.

Fewer errors are evident in the figures representing hit confirmation of the Asinex compounds, than for the other libraries. This is due to the fact that these hits were identified and confirmed using a refined assay that included 1mM KCl in the reaction buffer (see PUT and Figure 3.4.3C, above). Both the hit rate and false positive rate for this library were 0.7%.

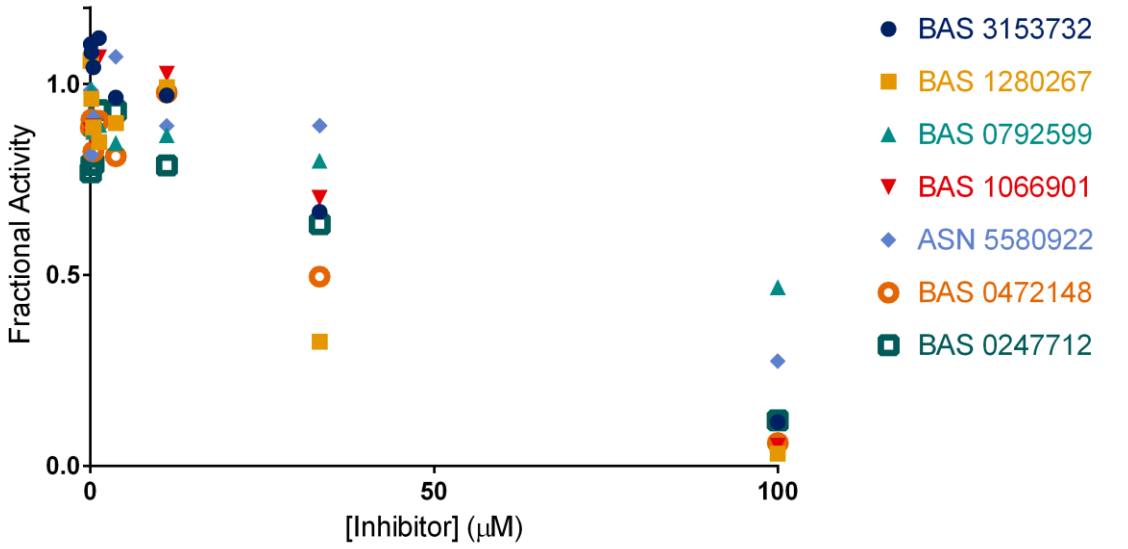


Figure 3.4.15 - plot of fractional activity against the concentration of all confirmed hits from the Asinex library.

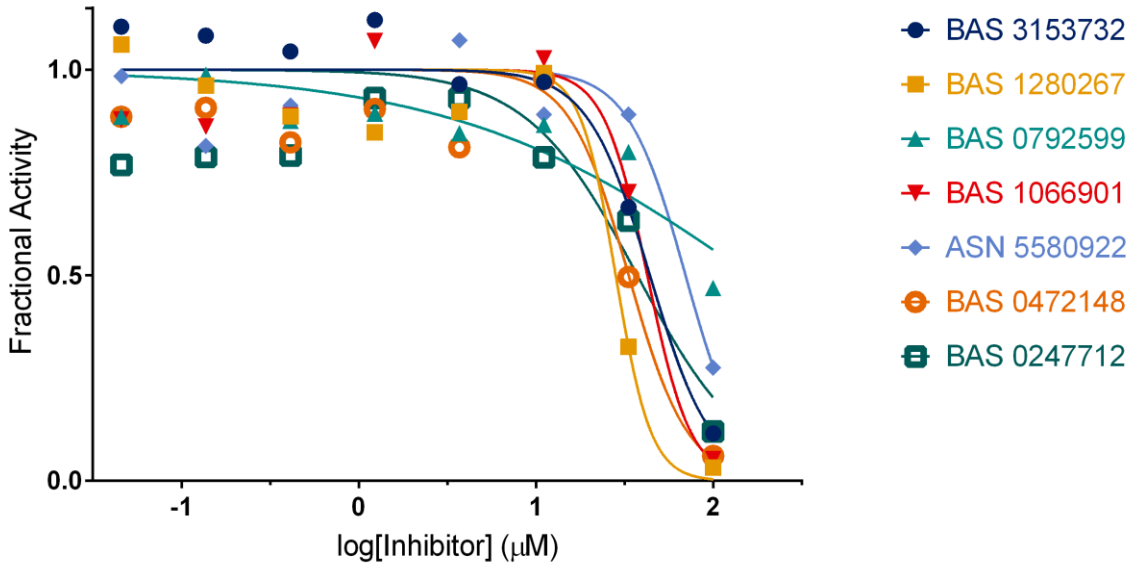


Figure 3.4.16 - plot of fractional activity against the log concentration of all confirmed hits from the Asinex library.

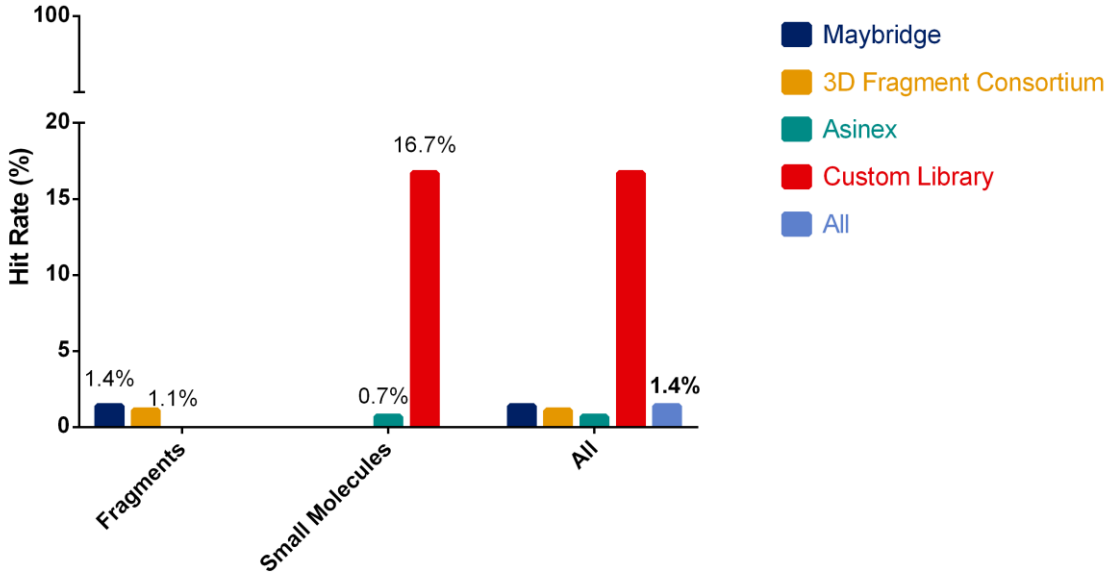


Figure 3.4.17 - hit rates of different subsets of the *in vitro* screening library and the entire library. The overall hit rate of the *in* vitro screening experiment was 1.4%

The overall hit rate for the *in vitro* screen is within expected ranges for a high-throughput screening assay(281). The hit rate for the custom library is particularly high and the reasons of this will be discussed further in this thesis.

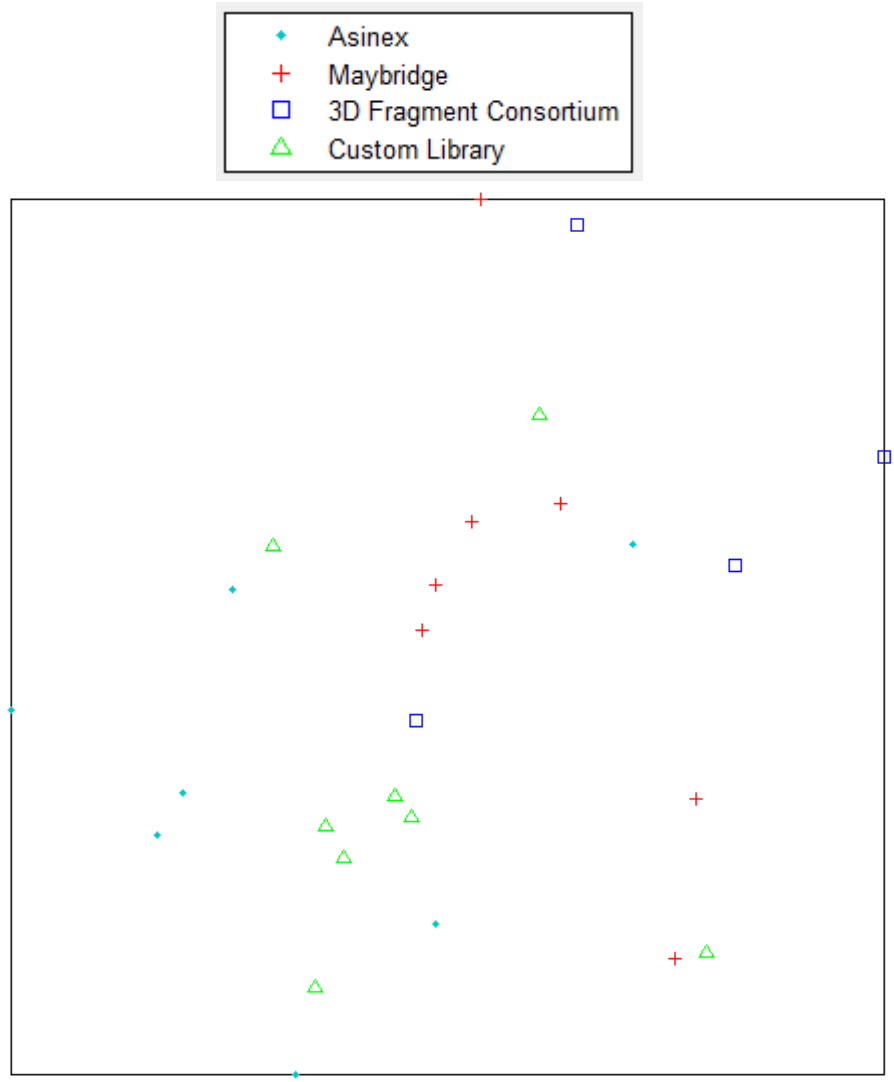
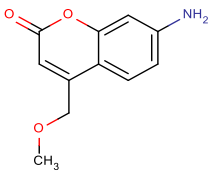
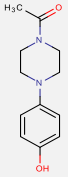
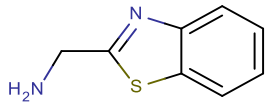
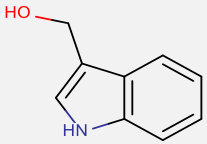
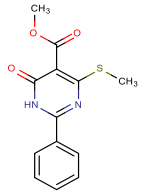
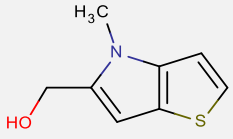
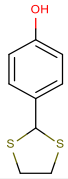
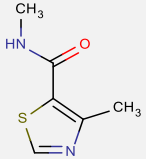


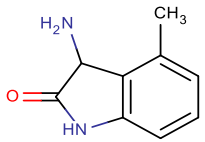
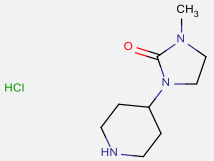
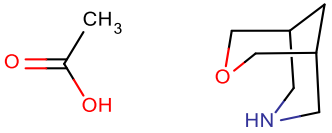
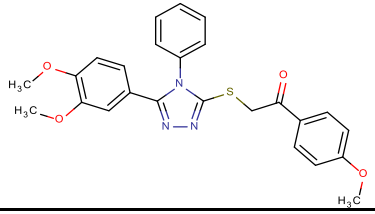
Figure 3.4.18 - visualisation of the diversity of all confirmed hits from *in vitro* screening. Each compound is represented by a different point, depending on which subset of the screening library it belongs. The relative positions of the points are determined by the differences between the chemical structures of the compounds, as in Figure 3.4.1.

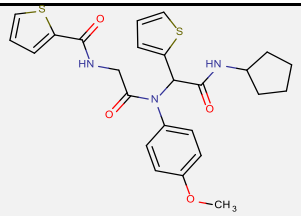
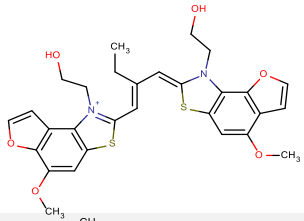
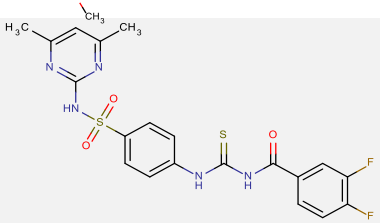
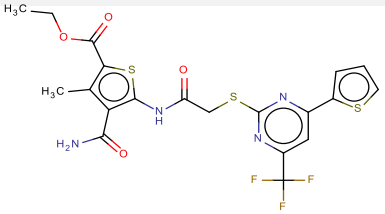
Given the modest library size, compared to libraries used in high-throughput screening, the screen was successful in identifying inhibitors with diverse chemical structures. The combination of various approaches, including the identification of potential hits using virtual screening and the application of fragment-based hit discovery, creates more options for future structure-based drug design efforts. The confirmed hits from the *in vitro* screen are listed below in Table 3.4.2.

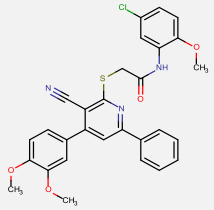
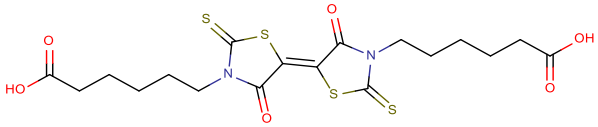
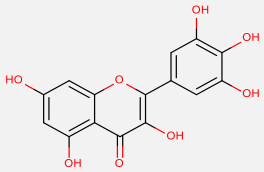
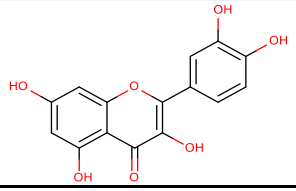
Table 3.4.2 - all confirmed hits and descriptive chemical terms from the *in vitro* screen.

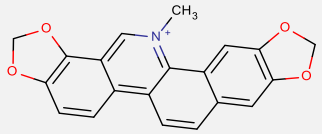
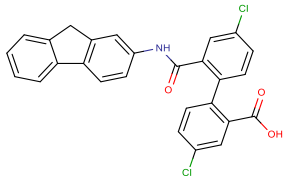
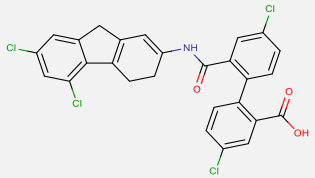
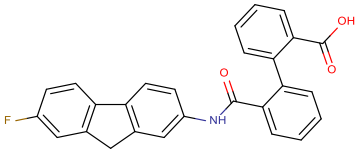
	Name	Structure	MW	LogP	TPSA	Hydrogen-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Congreve Rule of Three Compliant?
Fragments									
Maybridge	RJF 01106		205.21	0.62	61.55	3 , 1	2	✓	
	SB 02047		220.27	0.85	43.78	3 , 1	1	✓	✓
	CC 06013	HCl 	200.69	1.31	38.91	2 , 1	1	✓	✓
	CC 12709		147.17	1.30	36.02	1 , 2	1	✓	✓

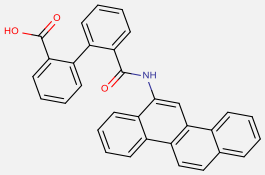
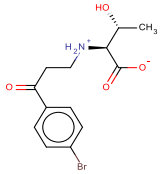
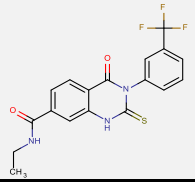
	Name	Structure	MW	LogP	TPSA	Hydrogen-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Congreve Rule of Three Compliant?
	BTB 08846		276.31	2.35	67.76	3 , 1	4	✓	
	CC 40909		167.23	1.33	25.16	1 , 1	1	✓	✓
	SPB 00826		198.31	2.74	20.23	1 , 1	1	✓	✓
3D Fragment Consortium	DDD00771052		156.21	-0.03	41.99	2 , 1	1	✓	✓

	Name	Structure	MW	LogP	TPSA	Hydrogen-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Congreve Rule of Three Compliant?
	4011502		162.19	0.76	55.12	2 , 2	0	✓	✓
	EN300-93218		219.71	-1.00	35.58	2 , 1	1	✓	✓
	EN300-95115		187.24	-0.24	21.26	2 , 1	0	✓	✓
Small Molecules									
Asinex	BAS 0792599		461.53	4.56	75.47	6 , 0	9	✓	

	Name	Structure	MW	LogP	TPSA	Hydrogen-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Congreve Rule of Three Compliant?
	ASN 5580922		497.63	3.68	87.74	4 , 2	9	✓	
	BAS 0472148		565.68	0.59	92.32	5 , 2	9		
	BAS 1066901		477.51	3.40	113.08	5 , 3	4	✓	
	BAS 3153732		530.56	5.48	124.27	5 , 2	10		

	Name	Structure	MW	LogP	TPSA	Hydrogen-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Congreve Rule of Three Compliant?
	BAS 1280267		546.04	6.25	93.47	6 , 1	9		
	BAS 0247712		490.64	3.40	115.22	6 , 2	12	✓	
Custom Library	myricetin		318.24	1.85	147.68	8 , 6	1		
	quercetin		302.24	2.16	127.45	7, 5	1	✓	

	Name	Structure	MW	LogP	TPSA	Hydrogen-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Congreve Rule of Three Compliant?
	sanguinarine		332.33	-0.94	40.8	4 , 0	0	✓	
	NSC 132249		474.34	7.34	66.4	3 , 2	4		
	NSC 132252		545.24	7.24	66.4	3 , 2	4		
	NSC 128598		423.44	6.28	66.4	3 , 2	4		

	Name	Structure	MW	LogP	TPSA	Hydrogen-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Congreve Rule of Three Compliant?
	NSC 128608		467.51	7.34	66.4	3 , 2	4		
	2-{[3-(4-bromophenyl)-3-oxopropyl]amino}-3-hydroxybutanoic acid (BPOB)		330.17	-1.01	94.04	4 , 2	7	✓	
	Qc1		393.38	3.83	61.44	2 , 2	4	✓	

3.4.4 Hit validation and characterisation

Hit validation is mainly centred on validating a hit's activity and its identity. In this study, hit validation involved verifying that the observed inhibition demonstrated by the hits was specific and not the result of interference with the assay. To this end, assays to accurately measure the IC_{50} and to determine the mode of inhibition (MOI) of the inhibitors were carried out. Tests for non-specific inhibition using detergent (Triton X-100) were also performed to exclude non-specific inhibitors (see methods).

3.4.4.1 Asinex Library

It was not possible to obtain a further supply of the hits from the Asinex library, so stock solutions from the original supply for screening were used in hit validation studies. To establish whether the Asinex inhibitors were acting by a specific mode of action, the hit confirmation assay was repeated using a reaction buffer containing 0.01% Triton X-100. The results are plotted in Figure 3.4.19 and Figure 3.4.20.

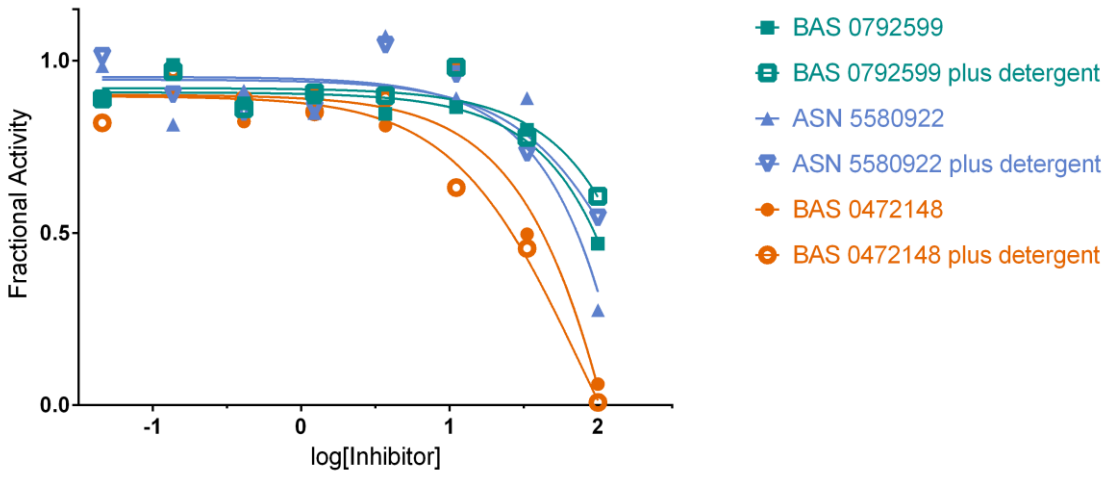


Figure 3.4.19 - plot of fractional activity against log inhibitor concentration. Confirmed hits from the Asinex library were tested in the presence and absence of detergent (Triton X-100). The inhibitors in this plot all showed similar levels of activity under both conditions.

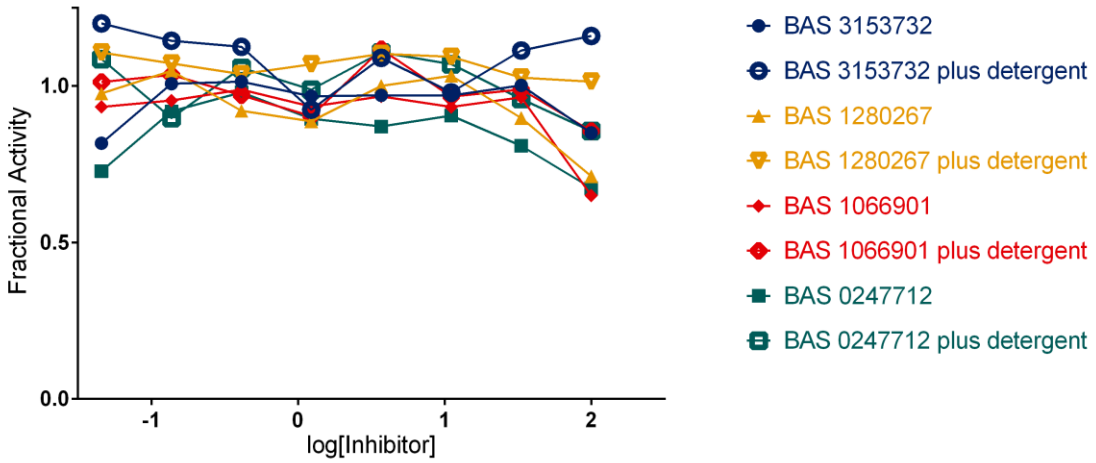


Figure 3.4.20 - plot of fractional activity against log inhibitor concentration. Confirmed hits from the Asinex library were tested in the presence and absence of detergent (Triton X-100). The inhibitors shown above all lacked activity in under both conditions.

Only three of the compounds were able to show inhibition as before. However, as these studies were carried out some months after the primary assay, the hit confirmation assay was also repeated without detergent. It was evident that four compounds which had previously shown inhibition were now showing considerably less inhibition (around 90%) or none at all. There was the possibility that these inhibitors were time-dependent, so the 30-minute pre-incubation time was insufficient (in the primary assay, inhibitors were pre-incubated with TDH for at least 30 minutes). The four inhibitors that had seemingly lost activity were re-tested at 100µM, after pre-incubation with TDH for at least 72 minutes, with and without detergent. After this, just one inhibitor, BAS 1066901, exhibited activity at a level close to the hit criteria in the primary assay: it inhibited TDH by 52.7% in the presence of detergent and 73.2% inhibition without detergent.

Table 3.4.3 – residual TDH activity after pre-incubation with Asinex inhibitors for at least 72 minutes, with or without detergent. Only BAS 1066901 shows significant inhibition.

	Percentage residual TDH activity	
Asinex ID	With detergent	Without detergent
BAS 3153732	92.8% (± 2.8%)	91.7% (± 7.2%)
BAS 1280267	85.5% (± 2.6%)	112.1% (± 8.8%)
BAS 1066901	52.7% (± 1.6%)	73.2% (± 5.8%)
BAS 0247712	76.8% (± 2.3%)	99.4% (± 7.8%)

Based on the results available at present, only four inhibitors can be confirmed: BAS 0792599, ASN 5580922, BAS 0472148 and BAS 10660901. IC_{50} values were calculated from the results of the first hit confirmation assays as 31.8µM for BAS 0792599, 84.3 µM for ASN 5580922, 33.3µM for BAS 0472148 and 36.1µM for BAS 1066901.

3.4.4.2 Fragment Library

Maybridge

Assays to determine the IC_{50} values of the fragments were carried out at the same substrate concentrations as in the primary screening assay.

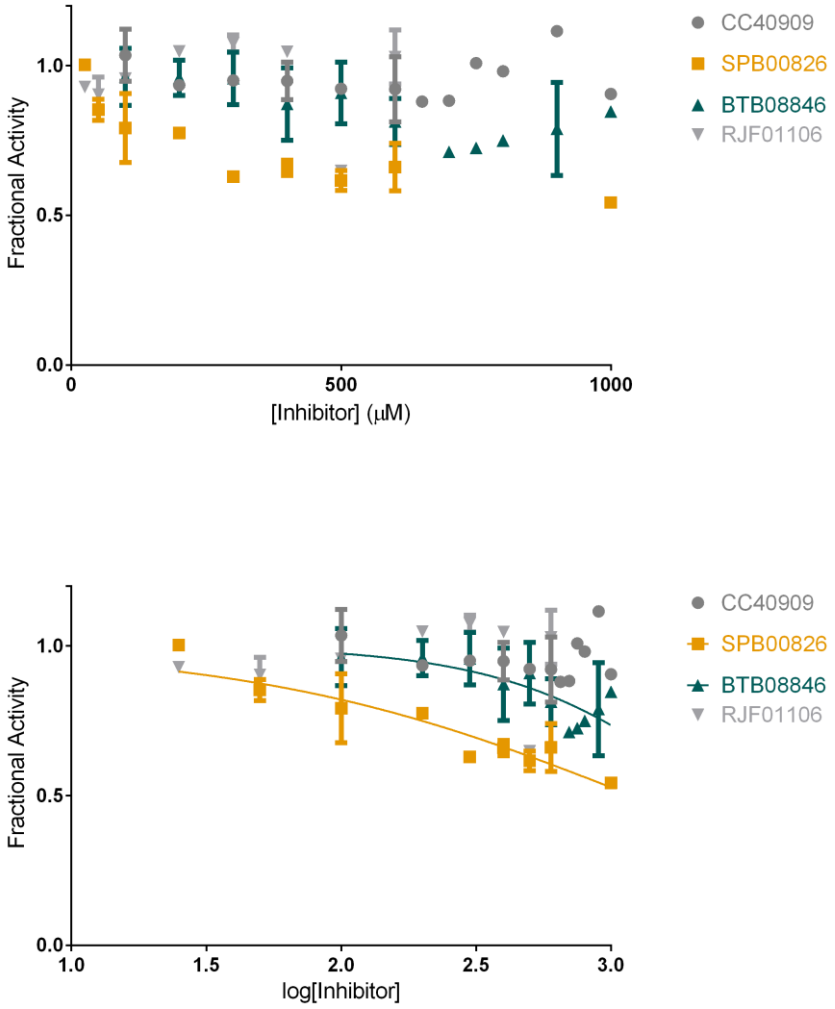


Figure 3.4.21 - plots of fractional activity versus inhibitor concentration (top) and fractional activity versus log inhibitor concentration (bottom), for four previously confirmed hits from the Maybridge library.

Six of the seven confirmed hits from the Maybridge library were re-procured from various commercial vendors. When testing these compounds across a range of different concentrations, the compounds CC40909 and RJF01106 failed to show any inhibition matching that which was found in the primary screening assay. SPB00826 showed weak inhibition, which was similar to the inhibition displayed in the primary screening assay. The difference in fractional activity at 1000μM and 500μM was relatively small, suggesting that, this compound is a partial inhibitor of TDH. The same conclusion may be made for BTB08846, which showed

around 71-85% inhibition over the same concentration range. SPB00826 and BTB08846 exhibited 21.5% and 46.8% inhibition, respectively at 500µM in the hit confirmation assay. However, fractional activity values at a third of this inhibitor concentration were 88.8% for SPB00826 and 119.6% for BTB88046. Thus the results are not necessarily conflicting. The higher inhibition in the screening and hit confirmation assays may have been caused by differences in the concentrations of stock inhibitor solutions. By curve fitting a dose response curve in the data analysis program Prism (see lower plot in Figure 3.4.21), the IC_{50} values for these inhibitors were estimated at 1.20mM for SPB00826 and 2.46mM for BTB08846. If they are indeed partial inhibitors, then their maximum inhibition values are approximately 50% at 1mM, for SPB00826, and 75% at 1mM for BTB08846.

As stated earlier, two potent fragments were identified in the Maybridge library: SB02047 and CC06013. These inhibitors were active in the low micromolar range and the IC_{50} values were 4.6µM for SB02047 and 132.8µM for CC06013.

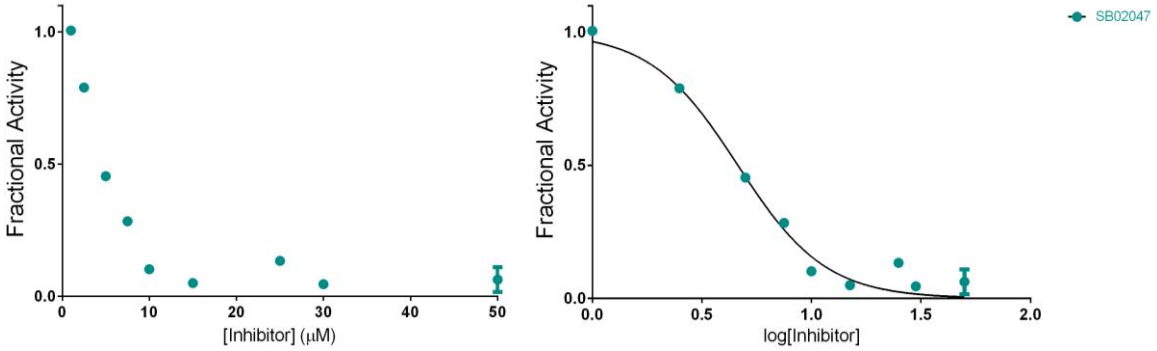


Figure 3.4.22 - plots of fractional activity as a function of SB02047 concentration (left) and log SB02047 concentration (right).

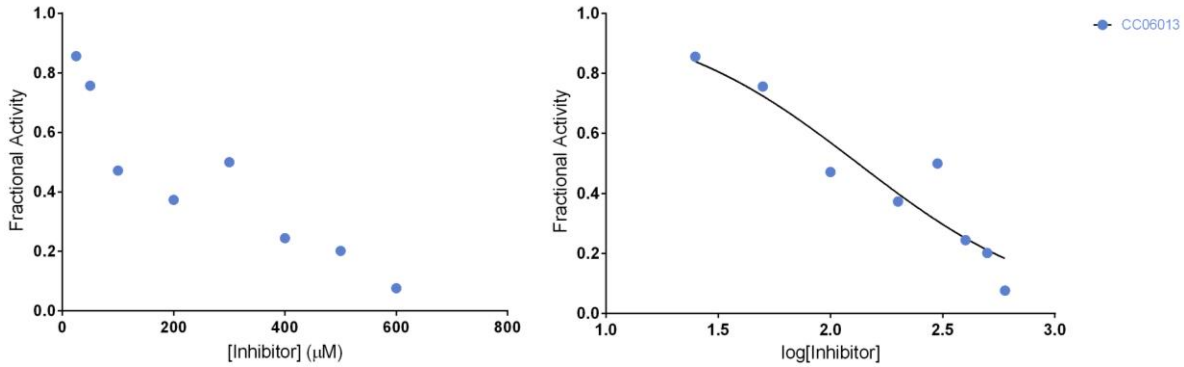


Figure 3.4.23 - plots of fractional activity as a function of CC06013 concentration (left), and log CC06013 concentration (right).

3D Fragment Consortium

The two stronger inhibitors from the four confirmed hits from the 3D fragment library were synthesized by Dr Sally Oxenford: DDD00771052, 4011502. A range of concentrations were tested against TDH to establish the IC_{50} concentrations. The results are plotted in Figure 3.4.24, below.

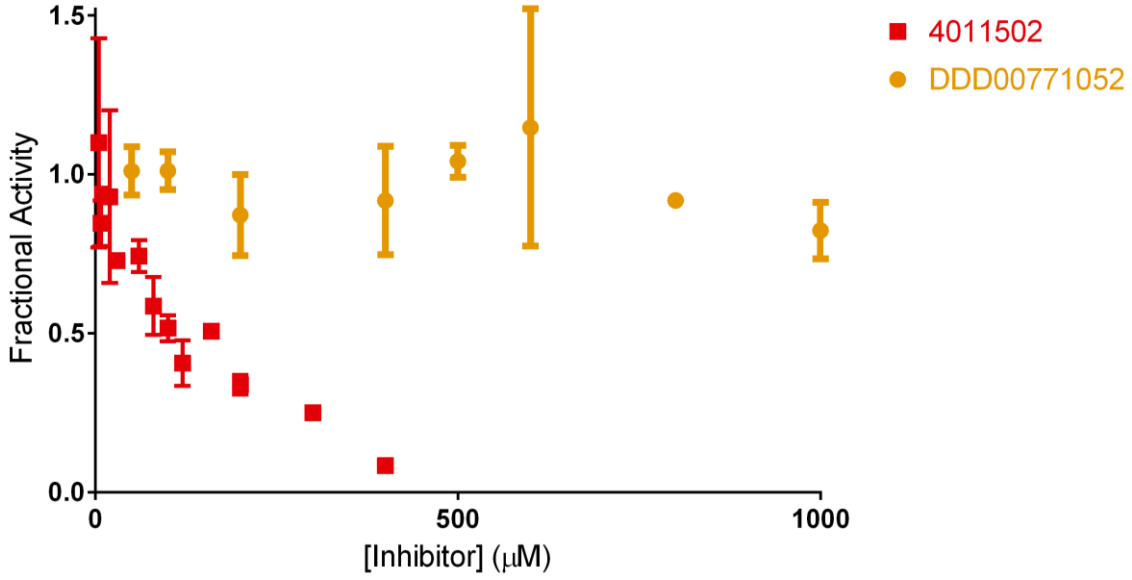


Figure 3.4.24 - plot of fractional activity against the concentration of fragments DD00771052 and 4011502. DDD00771052 does not show significant inhibition across the range tested, whilst 4011502 reproduces the inhibition demonstrated in hit confirmation assays.

DDD00771052 failed the validation assay and only demonstrated inhibition of up to 82% at 1mM. Conversely, 4011502 was able to demonstrate inhibition with similar potency to that seen in the primary and hit confirmation assays. An IC_{50} of 111.8μM was established for this compound.

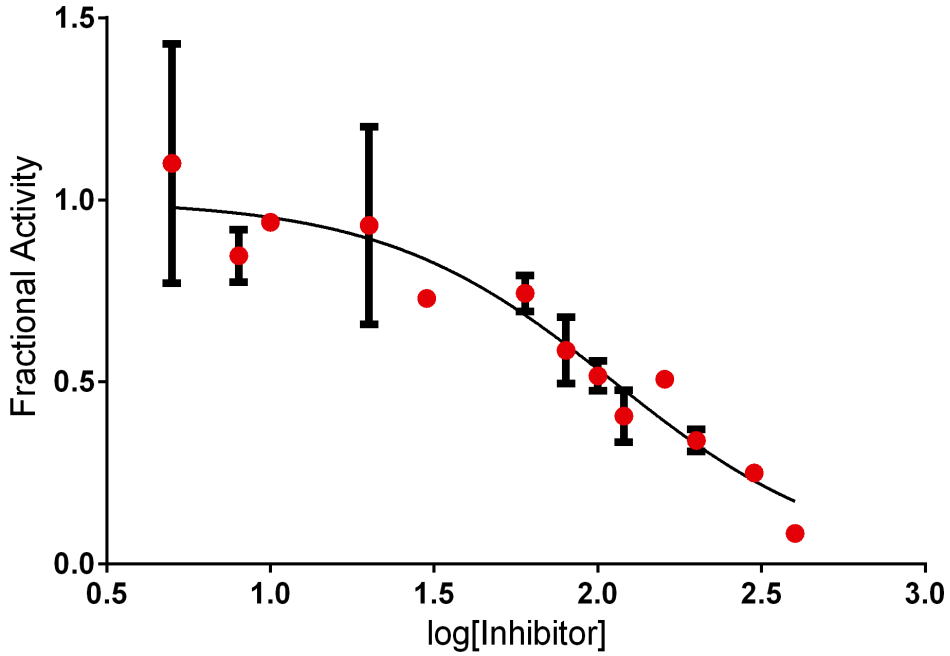


Figure 3.4.25 - plot of fractional activity versus log 4011502 concentration. A dose-response curve was fit to the data and an IC_{50} of 111.8µM was determined.

As with the Asinex library, the fragments were tested in the presence of 0.01% Triton X-100 to rule out non-specific inhibition by self-aggregation. The results of this test are listed in Table 3.4.4, below.

Table 3.4.4 - the effect on TDH activity of confirmed fragment inhibitors, with or without detergent (0.01% v/v Triton X-100) in the reaction mixture. The standard deviations of replicate measurements are shown in brackets.

			Residual TDH activity (standard deviation)	
Compound Library	**ID**	**Concentration**	**Without detergent**	**With detergent**
Maybridge	SPB00826	500µM	44.7% (± 7.3%)	56.4% (± 12.4%)
	SB02047	20µM	51.1% (± 7.7%)	47.5% (n.d.)
	CC06013	100µM	45.9% (± 6.1%)	43.9% (± 14.4%)
	BTB08846	500µM	87.8% (n.d.)	89.9% (n.d.)
3D Fragment Consortium	4011502	100µM	61.9% (± 7.0%)	48.1% (± 26.5%)

All of the inhibitors in Table 3.4.4 have undergone confirmation as specific inhibitors of TDH and 4011502 from the 3D Fragment Consortium is the most fully validated compound, having been synthesised, purified and characterised prior to testing.

3.4.4.3 Custom Library

Table 3.4.5 shows the inhibition of TDH with fixed concentrations of inhibitors from the custom library, in the presence and absence of 0.01% v/v Triton X-100.

Table 3.4.5 - the effect on TDH activity of confirmed inhibitors from the Custom library and TETD as a positive control, with or without detergent (0.01% v/v Triton X-100) in the reaction mixture. The standard deviations of replicate measurements are displayed in brackets, where applicable. * - similar values of percentage activity with detergent were found for a range of concentrations of NSC132252 (250 – 25µM), prompting further investigation (see below).

			Percentage residual TDH activity	
Compound Library	**ID**	**Concentration**	**Without detergent**	**With detergent**
Custom	Qc1	250µM	79.0%	65.0%
	sanguinarine	100µM	50.5% (± 0.4%)	57.6% (± 5.9%)
	myricetin	250µM	7.2%	5.4%
	quercetin	50µM	73.3% (± 4.4%)	79.4% (± 11.5%)
	NSC132249	100µM	73.3% (± 1.0%)	69.6% (± 2.4%)
	NSC128608	100µM	87.6%	84.0%
	NSC128598	125µM	71.4%	61.6%
	NSC132252*	25µM	3.0% (± 4.4%)	49.0% (± 5.7%)
	TETD	30µM	71.7% (± 6.7%)	65.4% (± 5.1 %)

Most of the Custom library inhibitors exhibited comparable levels of inhibition in the presence and absence of 0.01% v/v Triton X-100. The one exception was NSC132252, which showed inhibition with and without the detergent present, but inhibited TDH to very different extents under each condition. The observation that the inhibition by NSC132252 was similar across a range of different concentrations also suggested that it was acting in a time-dependent manner (during the assay lower concentrations were tested after higher concentrations, so the lower concentrations of inhibitor had longer pre-incubation times). To test this hypothesis, NSC132252 was incubated with TDH at a fixed concentration, with or without Triton X-100 and for different lengths of time before activity was measured at saturating substrate concentrations. The results of this test are shown in Figure 3.4.26.

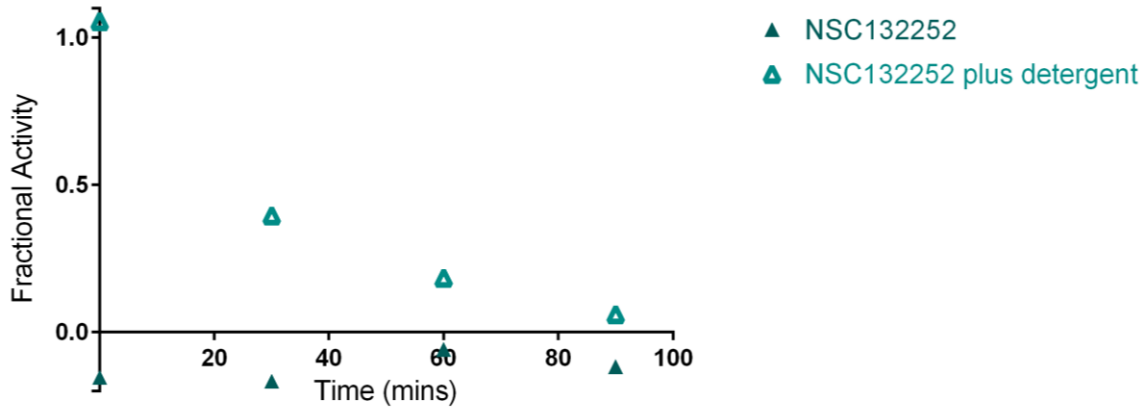


Figure 3.4.26 - the fractional activity of TDH following different incubation times with NSC132252, in the presence or absence of detergent. It is clear that inhibition is instant, in the absence of detergent, whereas inhibition is time-dependent in the presence of detergent.

Figure 3.4.26 clearly shows that, rather than abolish activity, as would be expected for a non-specific aggregating inhibitor, the addition of detergent appears to convert the activity of NSC132252 from that of a rapidly binding inhibitor to a time-dependent, slow binding inhibitor. There are a number of situations that could explain this. Firstly, the compound may inhibit TDH non-specifically or interfere with the assay readout in the absence of detergent, and the rapid inhibition exhibited is the composite of both the non-specific and specific time-dependent activity. An alternative explanation may be that NSC132252 inhibits TDH by two separate MOIs, one of which is interrupted by the presence of detergent. IC_{50} values and information on the compound's MOI were sought in the presence and absence of detergent to gain more information surrounding this inhibitor. The results of these studies will be discussed below.

In addition to NSC132252, several other compounds displayed time-dependent activity under normal conditions: BPOB, myricetin and the UDP-GalE inhibitors, NSC132249, NSC128598 and NSC128608.

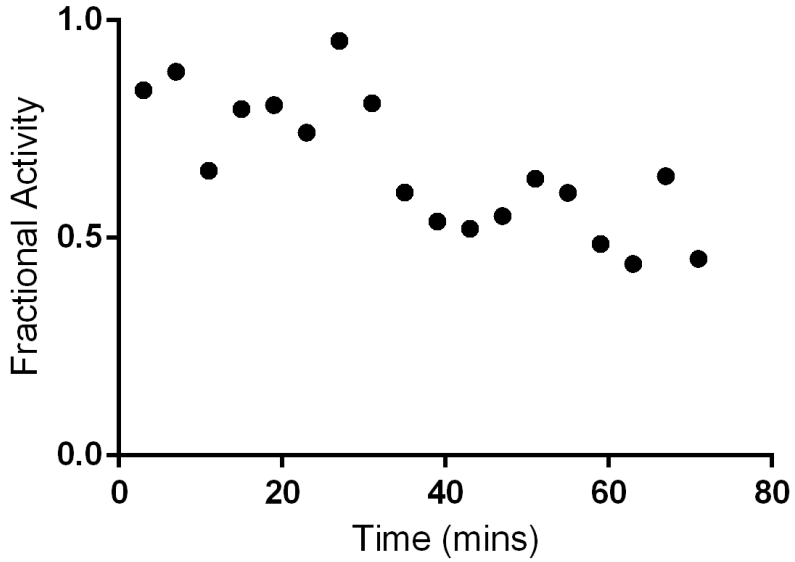


Figure 3.4.27 - time dependent inhibition of TDH by 1mM BPOB.

BPOB demonstrated a time-dependent inhibition that manifest in a linearly decreasing rate of reaction with time. At a concentration of 1mM, maximal inhibition had not yet been reached after 90 minutes of pre-incubation. For this reason, during MOI studies with time-dependent inhibitors, TDH was pre-incubated with the compound of interest for at least 2 hours prior to initiation of the reaction.

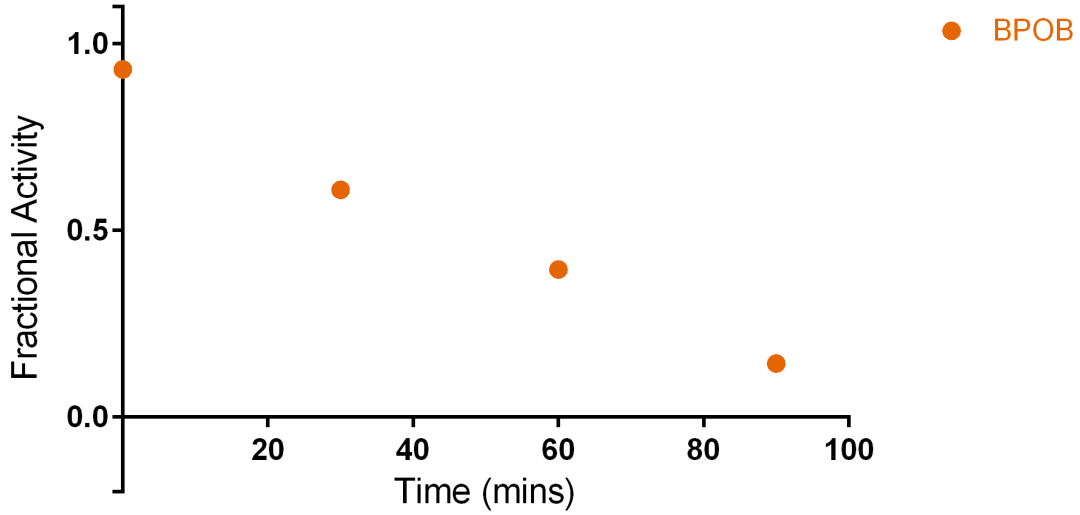


Figure 3.4.28 - time dependent inhibition of TDH by BPOB over pre-incubation times of up to 90 minutes.

The discovery of myricetin's time-dependent inhibition (see Figure 3.4.29, below), prompted the investigation of quercetin as a time-dependent inhibitor.

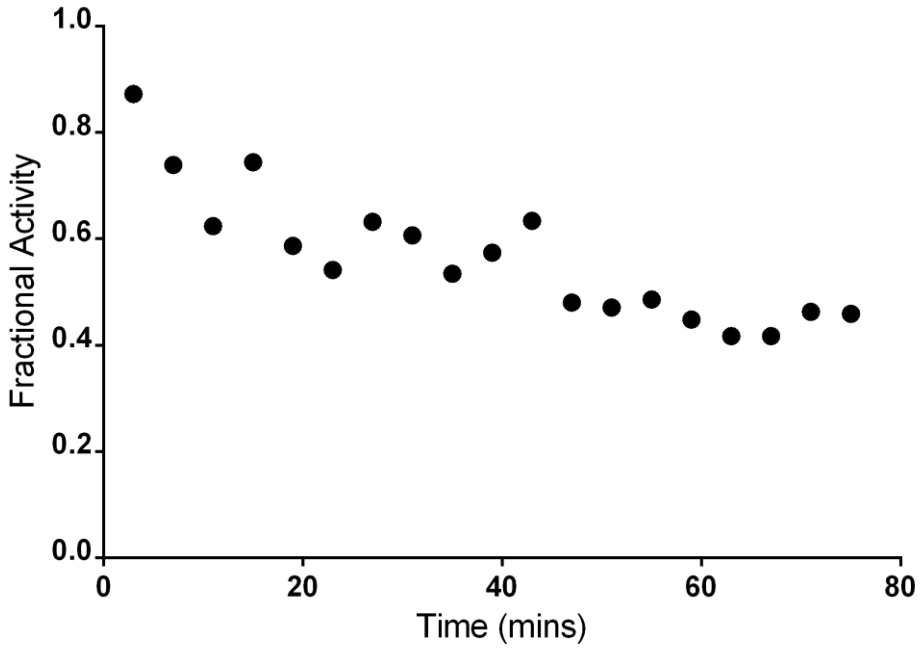


Figure 3.4.29 - time dependent inhibition of TDH by 100µM myricetin.

As quercetin is highly similar to myricetin, it was highly likely that it was also a time-dependent inhibitor. This was confirmed in an assay that measured TDH activity after pre-incubation times with quercetin of up to 110 minutes (see Figure 3.4.31).

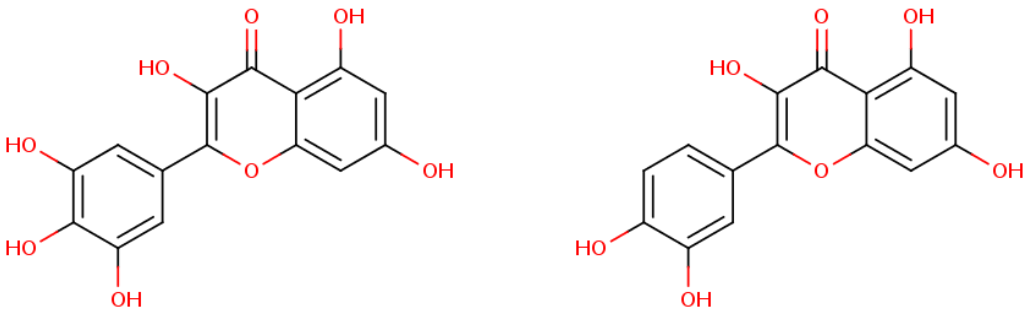


Figure 3.4.30 - the structural formulas of myricetin and quercetin. The two differ by just one hydroxyl group.

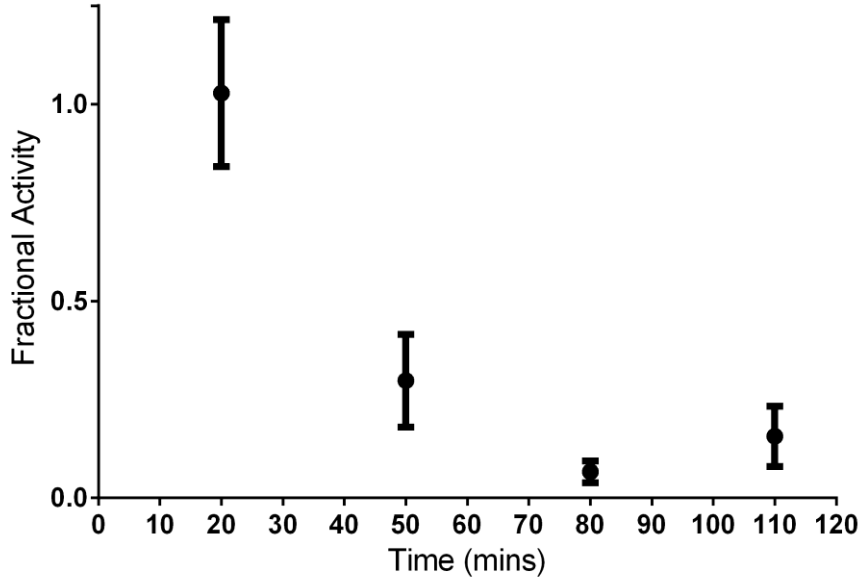


Figure 3.4.31 - time dependent inhibition of TDH by 300µM quercetin.

The UDP-GalE inhibitors, NSC132249 and NSC128598, caused a slow, linear decrease in TDH activity with increased pre-incubation time.

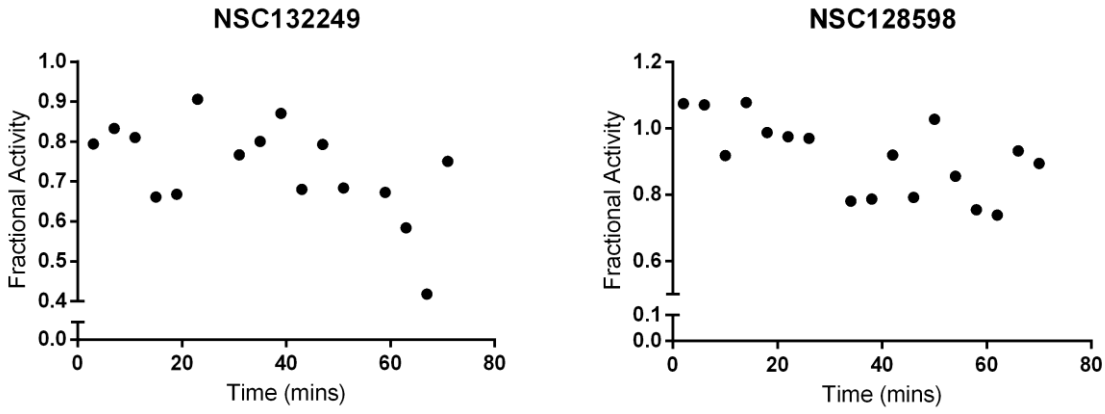


Figure 3.4.32 - time dependent inhibition by 500µM NSC132249 and 800µM NSC128598.

Taking this data and the data regarding NSC132252 into account, it was anticipated that this behaviour applied to the entire series of UDP-GalE inhibitors.

The assays to determine the IC_{50} values of the hits from the Custom library were carried out at saturating substrate concentrations so that the results could be used to inform the inhibitor concentrations to be used for MOI assays. As a consequence, the inhibitory concentrations presented here differ from those encountered in the primary screening and hit confirmation assays.

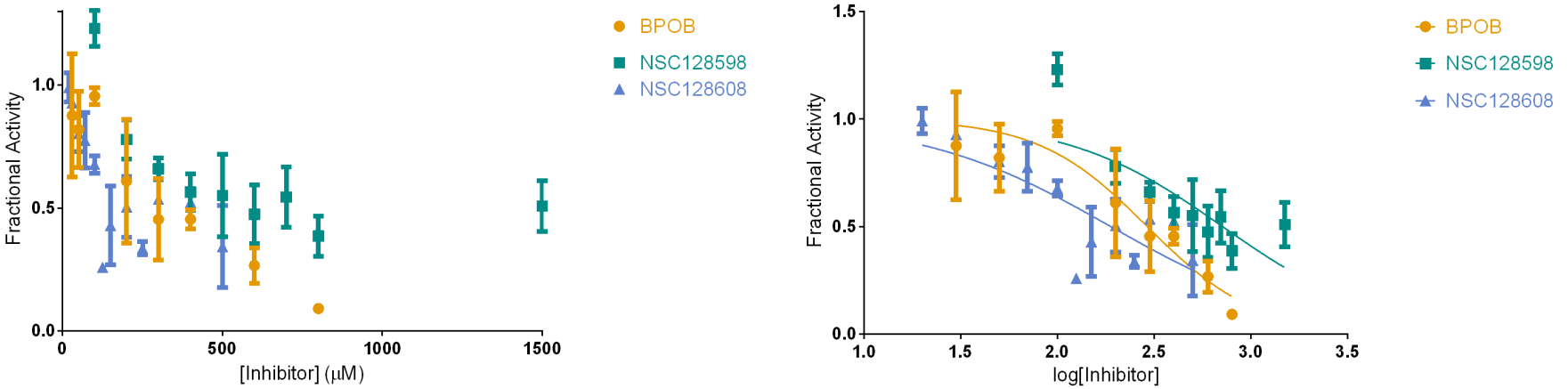


Figure 3.4.33 - fractional TDH activity as a function of the concentration of confirmed inhibitors from the Custom library. Dose-response curves were fitted to the data to calculate IC_{50} values of 291.4μM for BPOB, 713.9μM for NSC128598 and 188.7μM for NSC128608.

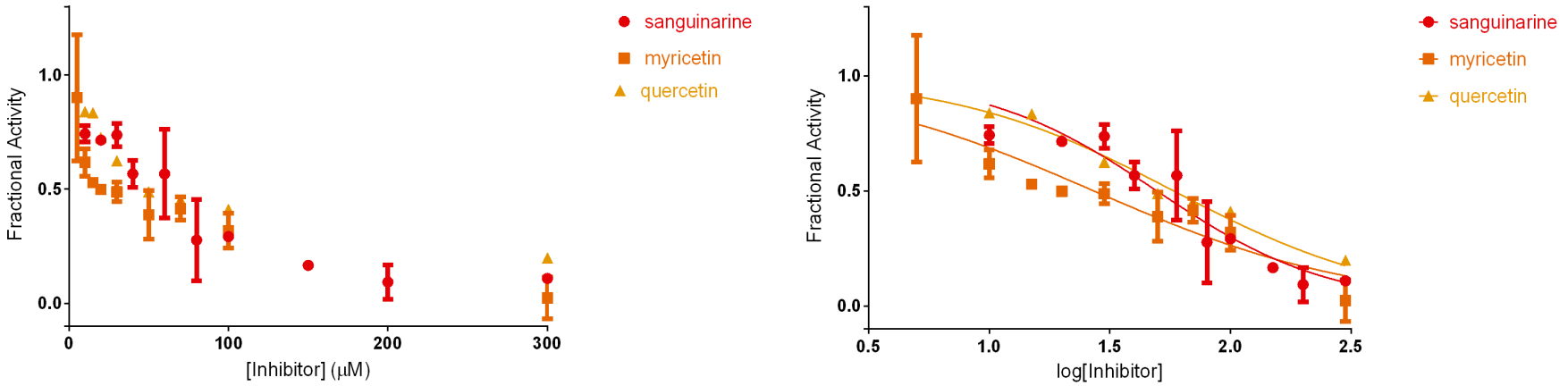


Figure 3.4.34 - fractional TDH activity as a function of the concentration of confirmed inhibitors from the Custom library. Dose-response curves were fitted to the data to calculate IC_{50} values of 49.55μM for sanguinarine, 27.15μM for myricetin and 58.21μM for quercetin.

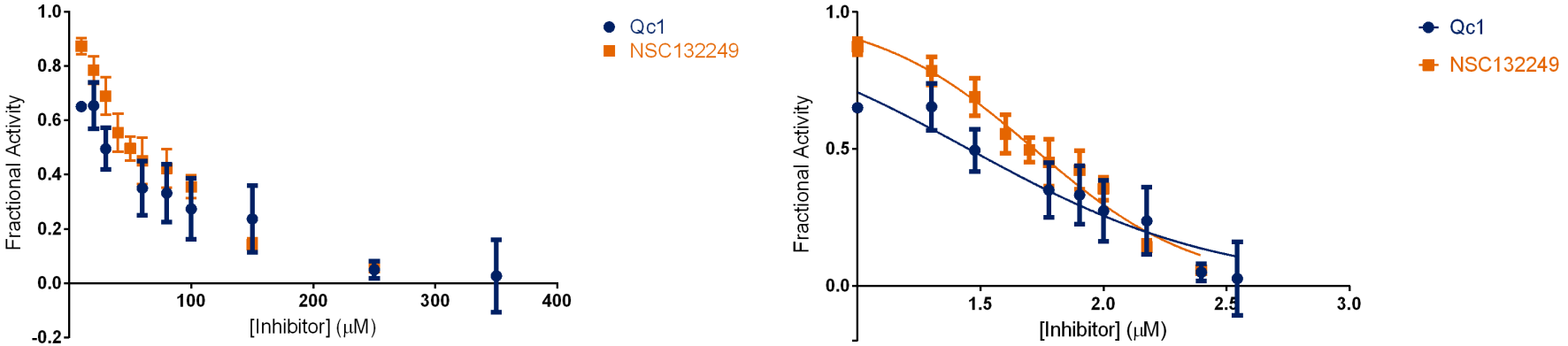


Figure 3.4.35 - fractional TDH activity as a function of the concentration of confirmed inhibitors from the Custom library. Dose-response curves were fitted to the data to calculate IC_{50} values of 28.3μM for Qc1 and 52.2μM for NSC132249.

The relative potencies of the hits were in agreement with those observed during hit confirmation assays. BPOB, NSC128598 and NSC128608 are the weaker inhibitors from this group of compounds. Qc1, sanguinarine, myricetin, quercetin and NSC132249 all had similar potencies, with IC_{50} values below 100µM. NSC132252 proved to be very potent, presenting an IC_{50} below 10µM. However, in the presence of detergent, its potency was comparable with that of NSC132249 (see Figure 3.4.35).

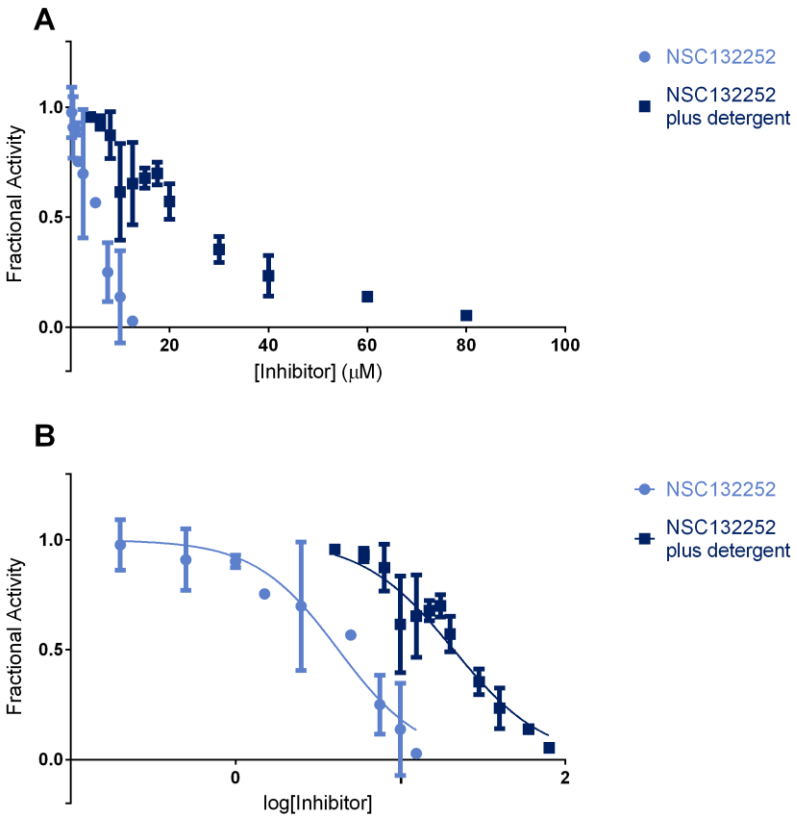


Figure 3.4.36 - fractional activity as a function of NSC132252 concentration (A) or log NSC132252 concentration (B) without detergent or with detergent. The potencies are very different under each condition. The IC_{50} under normal assay conditions is calculated as 4.2µM, whilst the IC_{50} with detergent is 21.0µM.

Following these results, detailed MOI studies were carried out on BPOB, Qc1, sanguinarine, myricetin, NSC132249 and NSC132252 (with and without detergent).

BPOB

The L-threonine analogue, BPOB is a time-dependent inhibitor. BPOB is one of the weaker compounds and had an IC_{50} of 291.4µM. The slope of a plot of rate as a function of BPOB concentration had a Hill coefficient of 1.54, suggesting that there may be some weak cooperativity involved in BPOB inhibition.

A curve based on the mixed model of inhibition was fitted to the data for fractional activity versus NAD^+ concentration with a high R^2 of 0.9180. The value of alpha was 2.15, but as this value was not significantly higher than the neutral value of 1.0 it suggested the presence of mixed competitive-noncompetitive inhibition (α indicates the affinity of the inhibitor for the free enzyme and the substrate-bound enzyme; values much greater than 1 indicate a preference for free enzyme, whilst values close to zero indicate a preference for substrate-bound enzyme). A competitive model fit to the data with a R^2 of 0.9161, but an analysis using Aikake's criterion (AIC) favoured the mixed model over the competitive model. Furthermore, the noncompetitive model (α = 1.0) fit to the data with a R^2 of 0.9177 and a comparison using AIC favoured this model over the mixed model. This suggests that the MOI of BPOB is indeed noncompetitive in relation to NAD^+. As there was some evidence of cooperativity in BPOB binding, the data were fitted using a Hill-modified mixed model, which gave similar results and an R^2 of 0.9184. The K_i was calculated as 821.1µM and a Hill coefficient of 1.33 was determined.

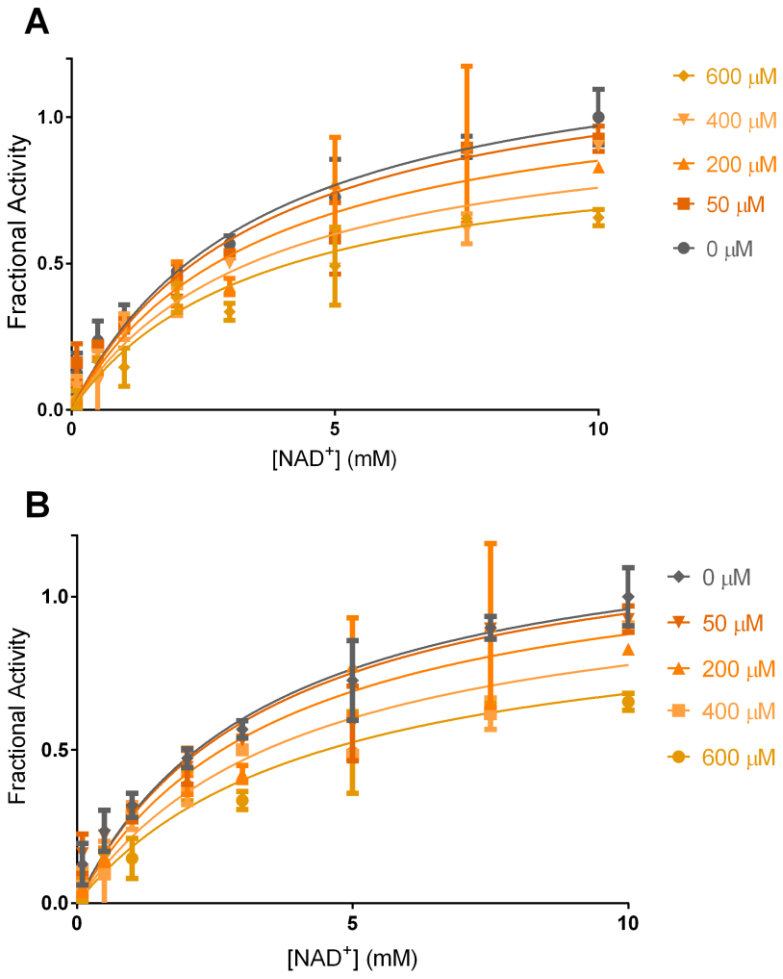


Figure 3.4.37 - plots of fractional activity versus NAD^+ concentration, at fixed concentrations of BPOB. The data are fit to curves with the equation describing mixed inhibition (A) and cooperative mixed inhibition (B).

As discussed earlier, as the relationship between L-threonine and rate does not follow Michaelis-Menten kinetics, the equations did not fit as well to the standard models. However, the relative fits of each model to the data was used to determine the most probable MOI. In this way, a noncompetitive MOI in relation to L-threonine was determined as the most likely MOI for BPOB. Both mixed and noncompetitive models fit with an R^2 of 0.808 and the mixed model fit with an alpha of 0.87. Fitting Hill-modified models to take account of cooperativity only brought marginal improvements to the fit to the data, and the Hill coefficient was small at 1.14.

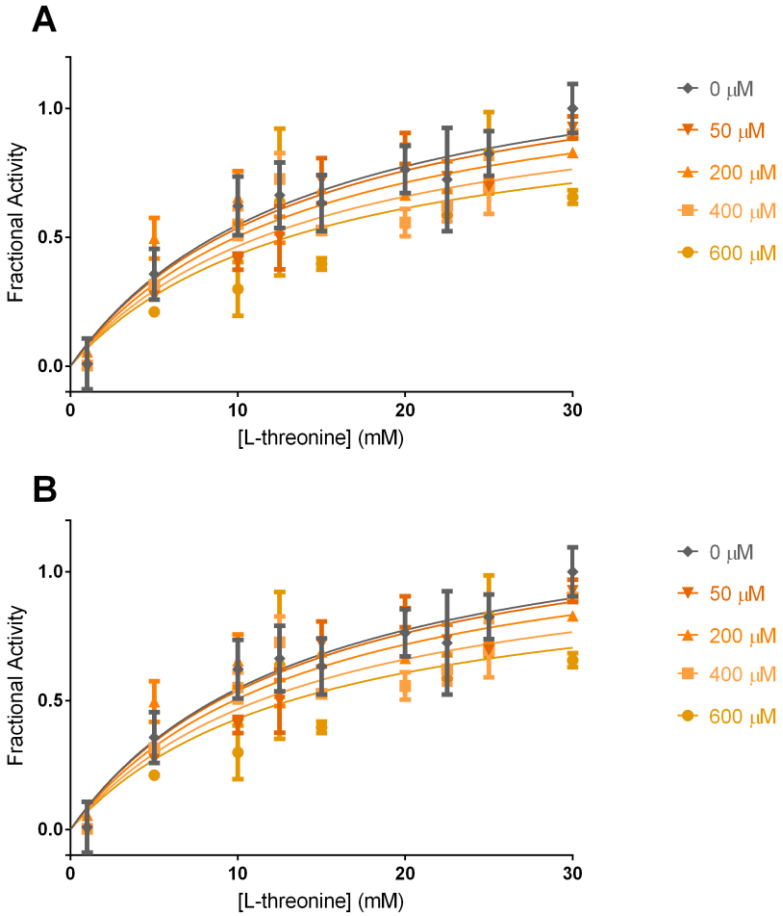


Figure 3.4.38 - plots of fractional activity versus L-threonine concentration, at fixed concentrations of BPOB. Curves were fitted using the equations describing noncompetitive inhibition (A) and cooperative noncompetitive inhibition (B). The fits of the curves to the data are almost identical, suggesting that cooperativity does not play a role in this respect.

With a noncompetitive MOI (when alpha=1.0), the K_i is equal to the IC_{50}. Using the IC_{50} stated previously, the K_i of BPOB can be estimated as 291.4µM.

These findings contrast with an MOI that one would expect for an L-threonine analogue. They also conflict with the binding modes predicted by Virtual Screening (see Figure 3.4.39 and Figure 3.4.40), which would have predicted an MOI that was competitive towards both substrates. Noncompetitive MOIs can sometimes indicate a nonspecific mechanism of action,

and this may help to explain why co-crystallisation with TDH failed. However, there was no other indication that the compound was acting non-specifically.

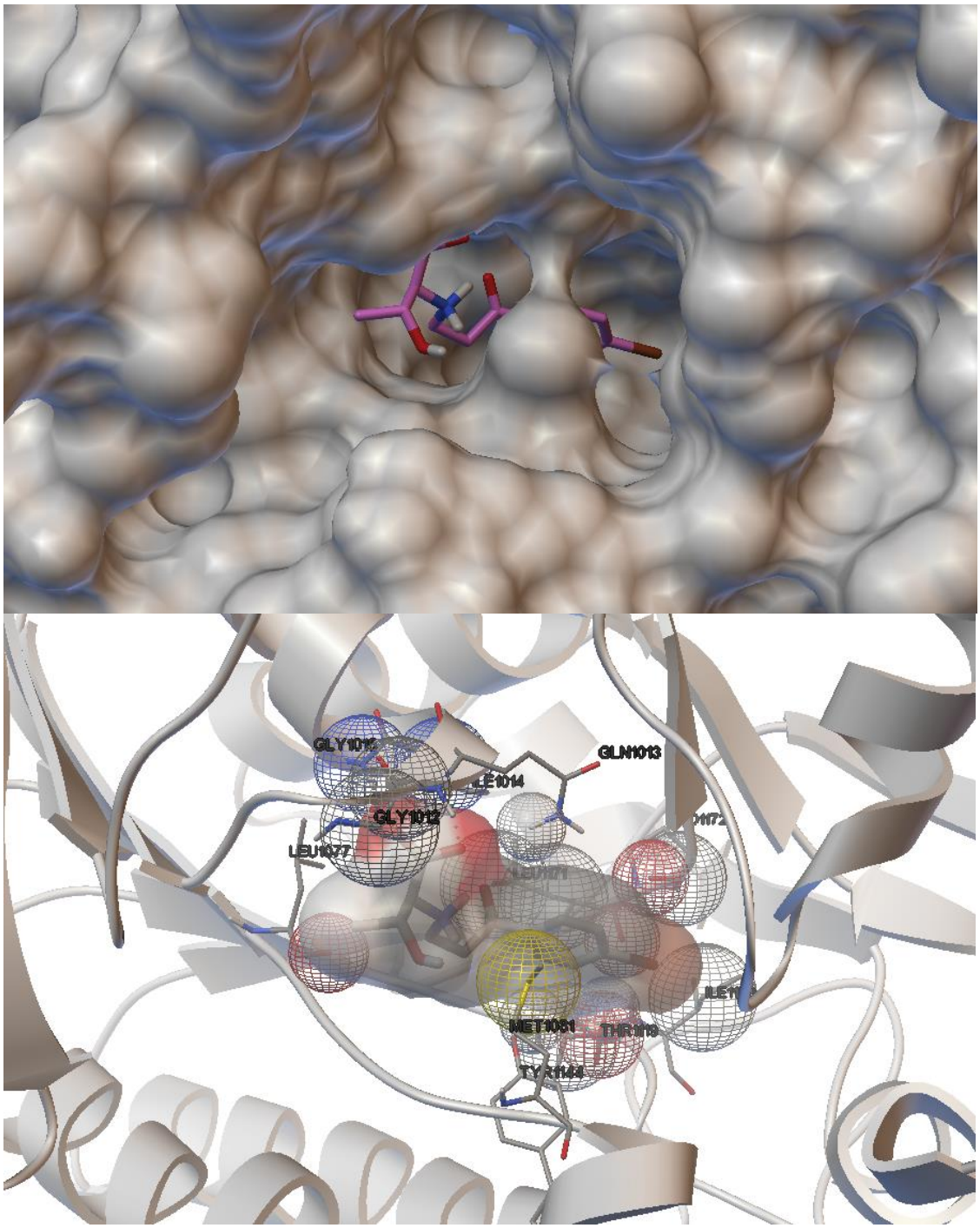


Figure 3.4.39 - binding mode of BPOB predicted in Virtual Screen 3. BPOB is predicted to bind both substrate sites, thereby competing with both substrates for binding. See also, Figure 3.3.15 and Figure 3.3.16.

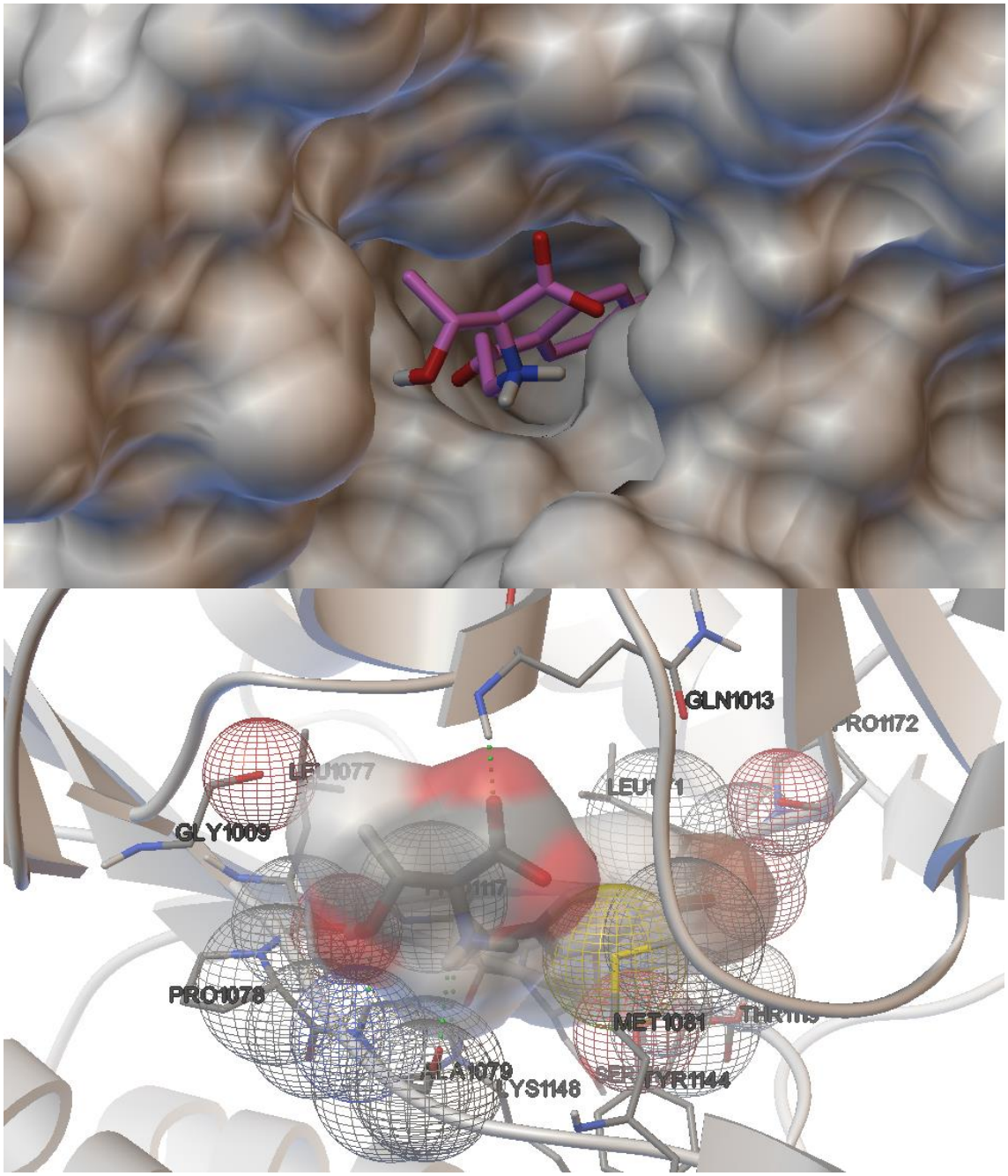


Figure 3.4.40 - binding mode of BPOB predicted in Virtual Screen 4. BPOB is predicted to bind both substrate sites, thereby competing with both substrates for binding. See also, Figure 1.2.15 and Figure 1.2.16.

Qc1

The IC_{50} of Qc1 was 28.3µM. A Hill coefficient close to unity (0.85), suggested that there was no cooperativity involved in its action. The results of data analysis of the relationship between NAD^+ and fractional activity indicated that Qc1 had an uncompetitive MOI in relation to NAD^+. The mixed model of inhibition fit with an R^2 of 0.8362 and yielded an alpha value of 0.278. The uncompetitive model fit with an R^2 of 0.8353 and determined an αK_i of 32.0µM. When this value is converted to K_i by division with the alpha value of 0.278, it results in a value that is very similar to that predicted by the mixed model – the K_i is 115.0µM, compared to 132µM, predicted

by the mixed model. These data suggest that Qc1 preferentially binds to the TDH-NAD^+ complex, but may also have low affinity for the *apo*-enzyme.

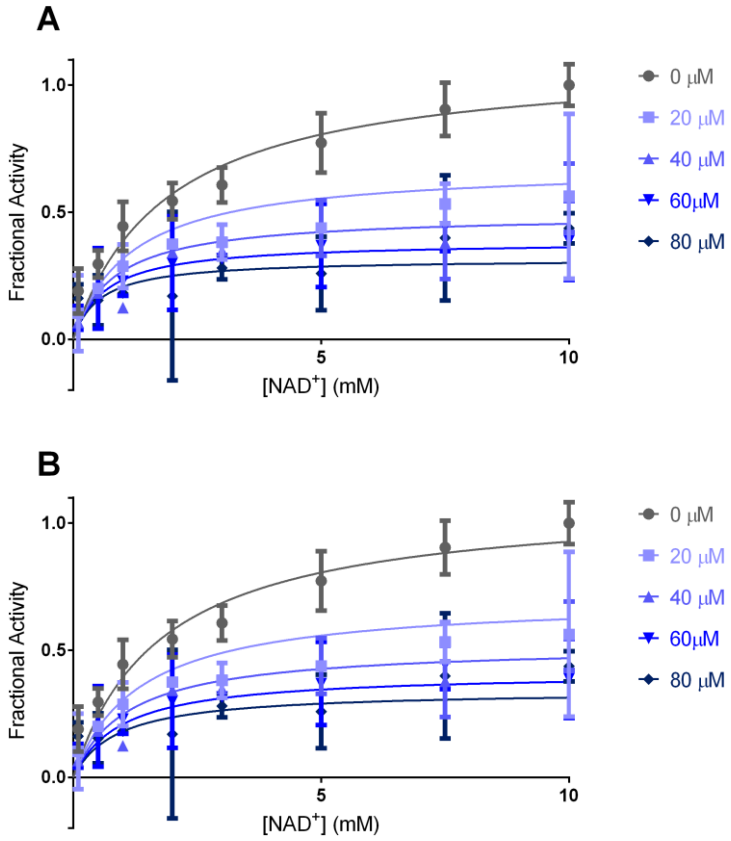


Figure 3.4.41 - plots of fractional activity versus NAD^+ concentration, at different fixed concentrations of Qc1. Curves describing mixed inhibition (A) and uncompetitive inhibition (B) are fitted to the data. It is evident that the fits to the data are very similar, suggesting that they represent the same or very similar MOIs.

The data obtained on the effect of Qc1 on the relationship between L-threonine concentration and rate of reaction varied significantly between replicate runs. However, the trends showing Qc1 concentration-dependent inhibition were the same. The MOI of Qc1 towards L-threonine was shown to be noncompetitive and the K_i was calculated as 46.7µM.

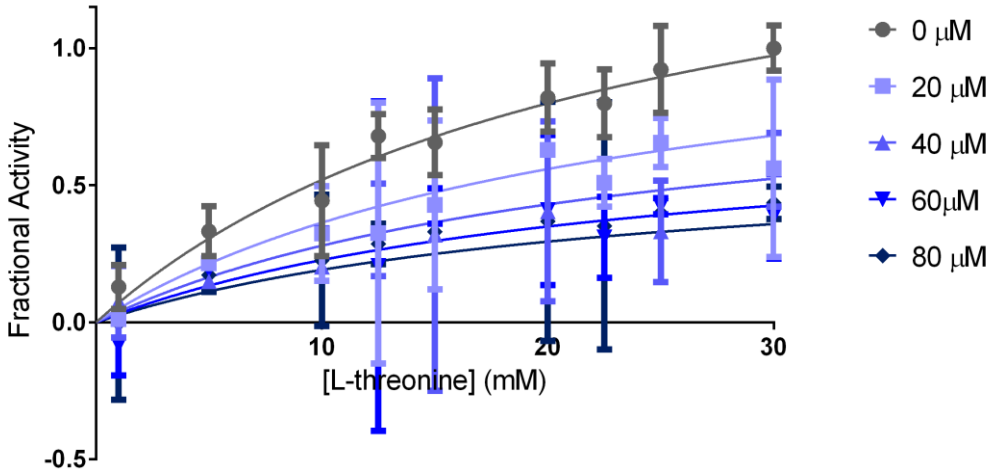


Figure 3.4.42 - plot of fractional activity against L-threonine concentration, at different fixed concentrations of Qc1. Curves describing noncompetitive inhibition are fit to the data.

The MOI implicated by the data analyses described above suggests that Qc1 may inhibit TDH by preferentially binding to the NAD^+-bound enzyme, but not by competing with L-threonine for binding. This is in disagreement with the binding modes predicted for Qc1 in Virtual Screens 3 and 4, in which other members of the Qc series were hits.

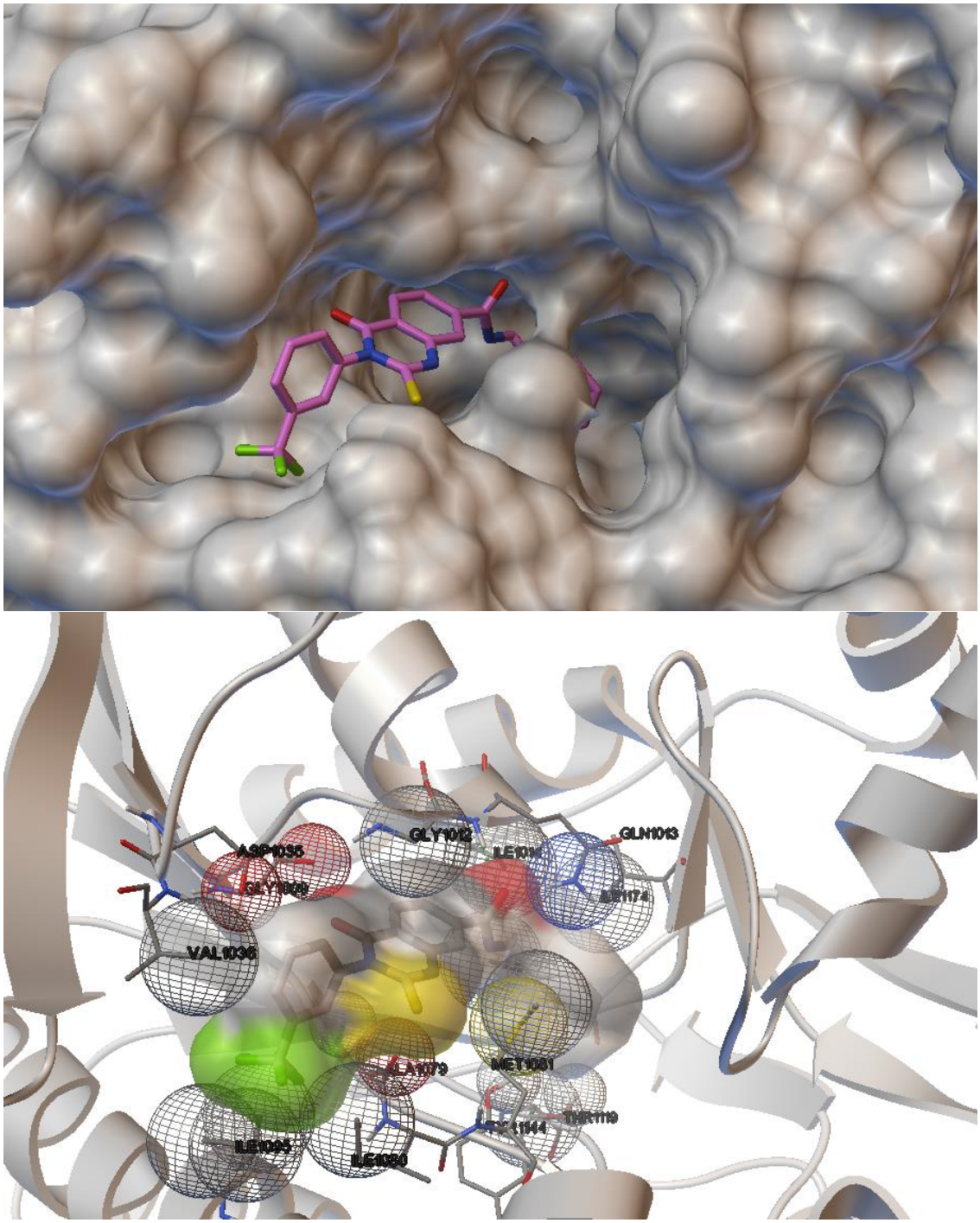


Figure 3.4.43 - predicted binding mode of Qc1 in Virtual Screen 3. The compound is predicted to bind in the NAD^+ binding site.

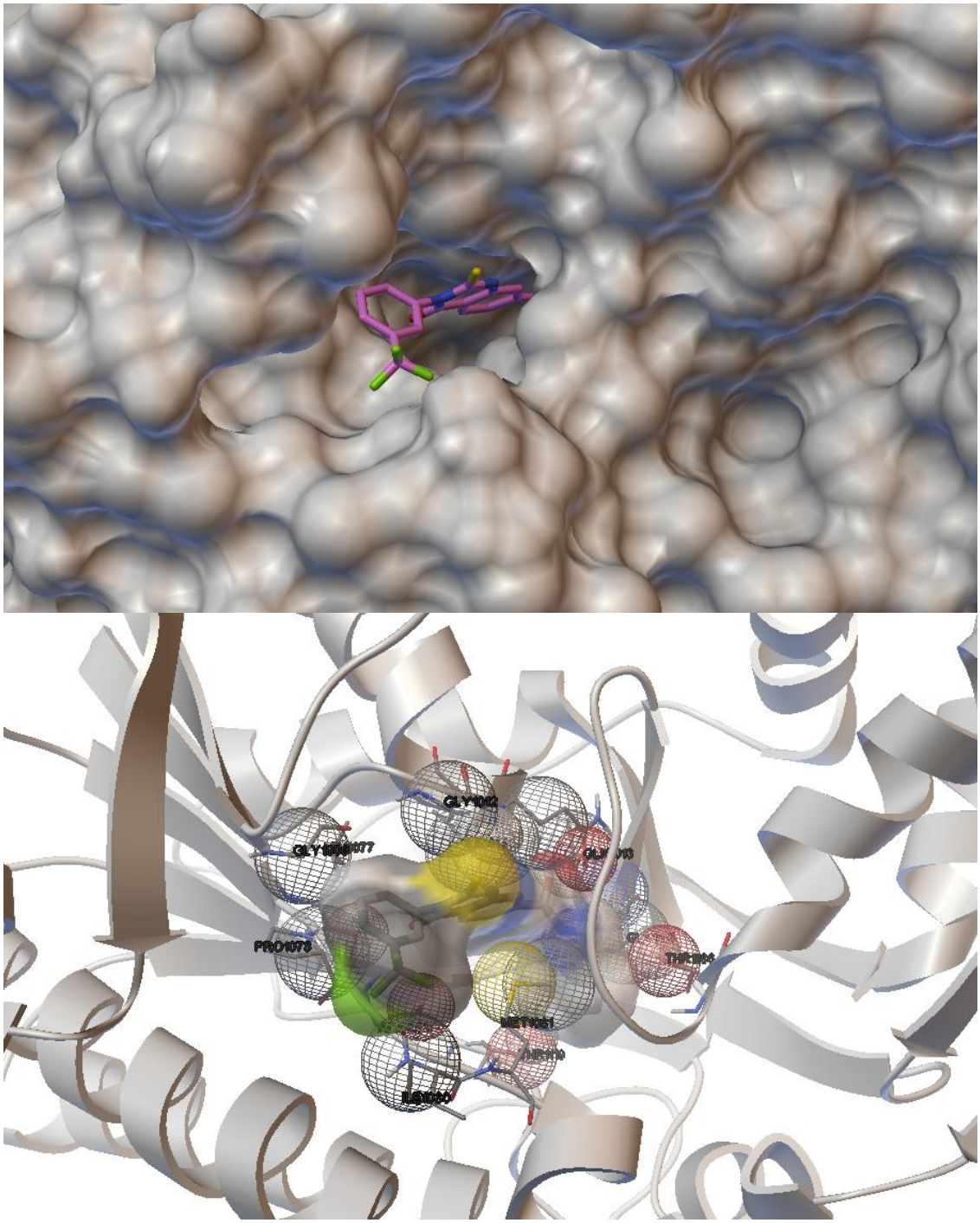


Figure 3.4.44 - predicted binding mode of Qc1 in Virtual Screen 4. Qc1 was predicted to bind in the NAD^+ binding site, as in Virtual Screen 3.

As with the L-threonine analogues, L-*allo*-threonine and pyruvate, one would expect an inhibitor with an MOI that was uncompetitive in regards to NAD^+ to bind the L-threonine binding pocket. On the contrary, the MOI towards L-threonine was noncompetitive. This raises the possibility that Qc1 may bind to a site other than the L-threonine and NAD^+ binding pockets.

Sanguinarine

The slope of the curve relating sanguinarine concentration with TDH fractional activity had a Hill coefficient of 1.2, suggesting that there was little or no cooperativity involved in the inhibition caused by sanguinarine. An IC_{50} of 49.6µM was calculated.

Analysis of data from enzymatic assays predicted that sanguinarine competes with NAD^+ for binding. MOI models for both competitive and mixed inhibition fit to the data with a R^2 of 0.874. The mixed model calculated a very high alpha value of approximately $2x10^{16}$, which strongly indicates a competitive MOI.

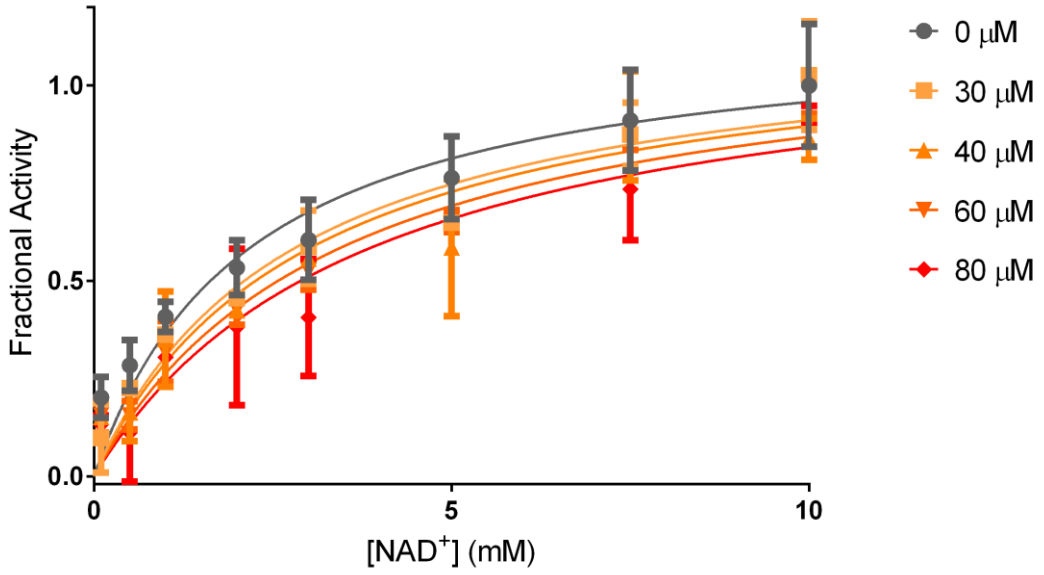


Figure 3.4.45 - plot of fractional activity versus NAD^+ concentration, at different fixed concentrations of sanguinarine. A curve describing competitive inhibition is fit to the data.

A competitive MOI in relation to L-threonine was also predicted by both mixed and competitive models, which fit with an R^2 of 0.7325.

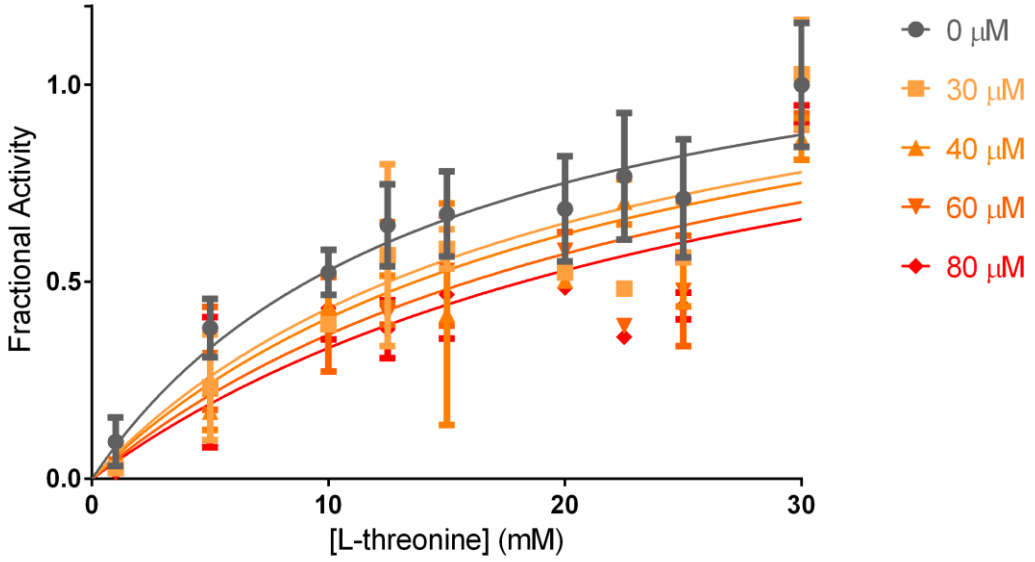


Figure 3.4.46 - plot of fractional activity versus L-threonine concentration, at different fixed concentrations of sanguinarine. Curves describing competitive inhibition are fit to the data.

With knowledge of the MOI of sanguinarine, the K_i could be estimated using its IC_{50}. Using the Cheng-Prusoff equation for competitive inhibitors, a K_i of 4.5μM was calculated. The MOIs demonstrated *in vitro* are in agreement with those predicted by virtual screening. Observing the predicted binding modes in Figure 3.4.47, one can see that in Virtual Screen 4, sanguinarine was predicted to bind in both the NAD^+ and L-threonine binding pockets.

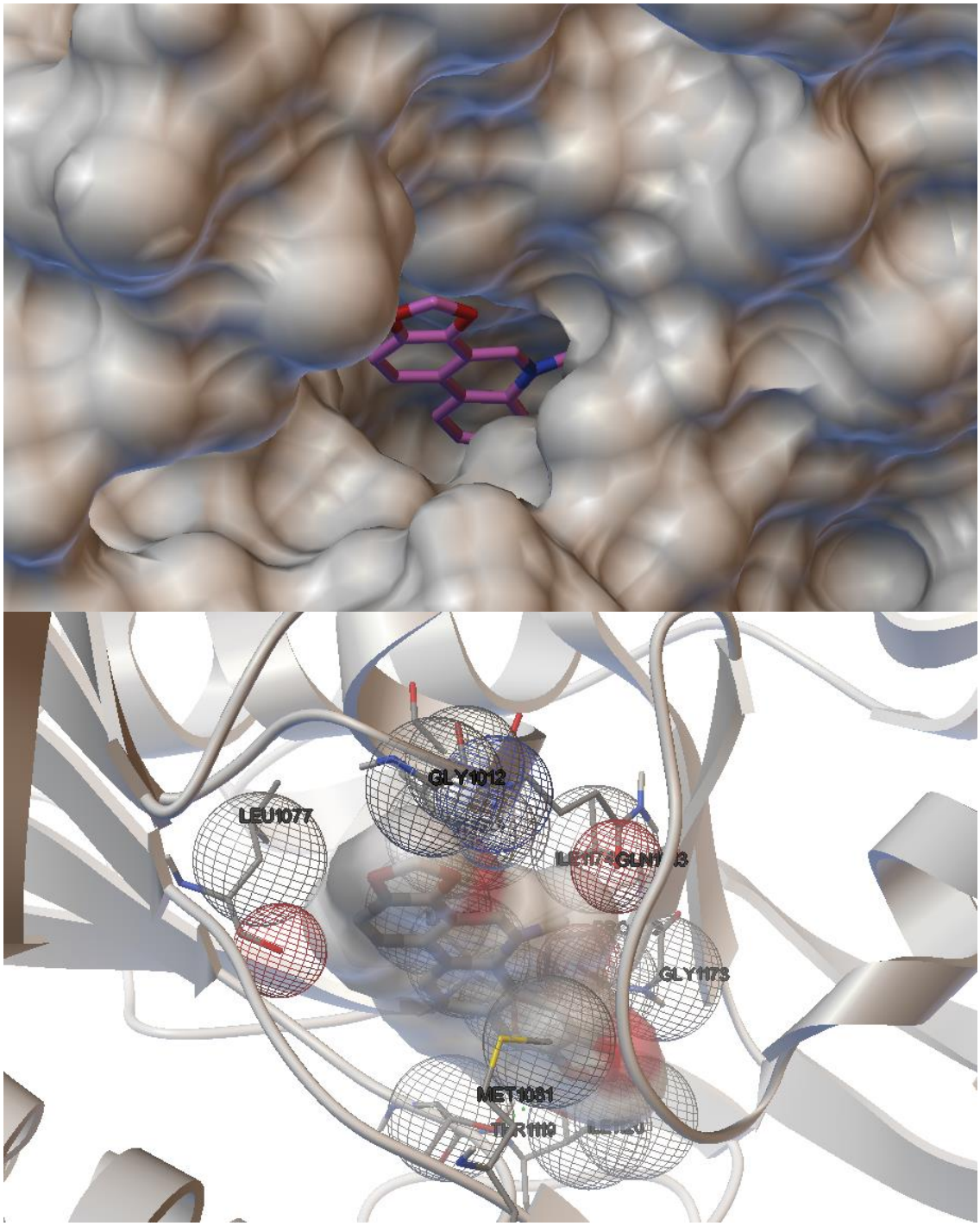


Figure 3.4.47 - binding mode of sanguinarine, predicted in Virtual Screen 4. It largely occupies the NAD^+ binding pocket, but also binds L-threonine binding residues Met81 and Thr119.

Myricetin

The inhibitor of dihydroflavonol 4-reductase (DFR), myricetin inhibited TDH potently with an IC_{50} of 27.2µM. The slope fit to the same data used to determine the IC_{50} had a Hill coefficient of 0.78, suggesting that there may be some negative cooperativity involved in myricetin binding.

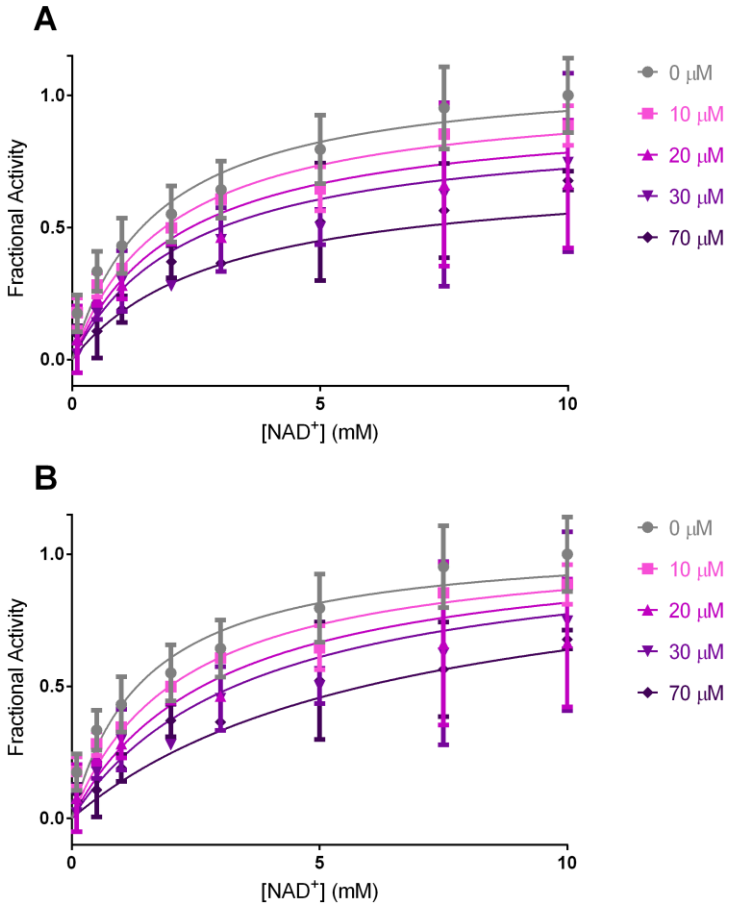


Figure 3.4.48 - fits of fractional activity versus NAD^+ concentration at fixed concentrations of myricetin. Curves describing mixed inhibition (A) and competitive inhibition (B) are fit to the data.

The best fitting model of inhibition describing myricetin's behaviour towards NAD^+ was the mixed model of inhibition. Alpha was predicted as 3.19, suggesting that the MOI was competitive or mixed noncompetitive-competitive. AIC comparisons of different models suggested that the mixed model was a better description of the data than the competitive model, and that a noncompetitive MOI was actually the most appropriate MOI to be assigned to myricetin. The K_i calculated by this model was 20.24µM.

It was clear from the data analyses that myricetin exhibited a competitive MOI against L-threonine. A competitive model calculated a K_i of 41.7µM. Fitting a Hill-modified competitive model resulted in slightly improved fits to the data, from an R^2 of 0.794 to 0.810. Consistent with the earlier evidence of negative cooperativity, a Hill-coefficient of 0.44 was calculated.

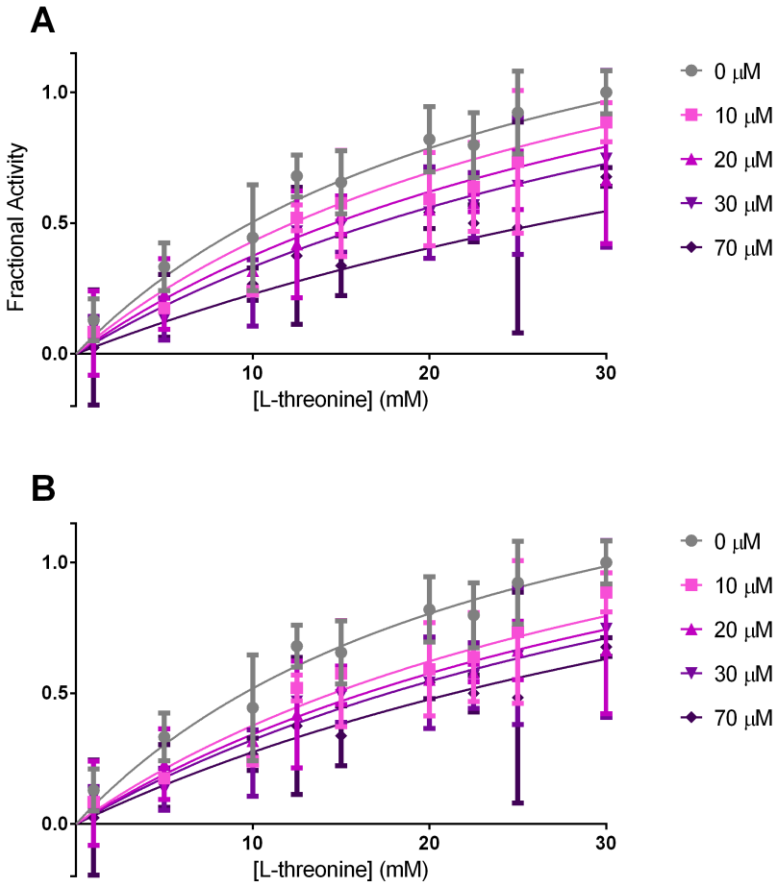


Figure 3.4.49 - plots of fractional activity as a function of L-threonine concentration, at different fixed concentrations of myricetin. Curves were fit with the competitive inhibition model (A) and the cooperative competitive inhibition (B) model.

All of the K_i values lie within a similar range, and the K_i that would be calculated by applying the Cheng-Prusoff equation for a noncompetitive inhibitor, equal to 27.2μM, is within the 95% confidence interval of the K_is calculated by non-linear regression. The conclusions drawn from these results suggest that myricetin may inhibit TDH by competing for the L-threonine binding site, in a way that is independent of NAD^+ binding. This MOI can be compared with that predicted during Virtual Screening by observing Figure 3.4.50 and Figure 3.4.51, below.

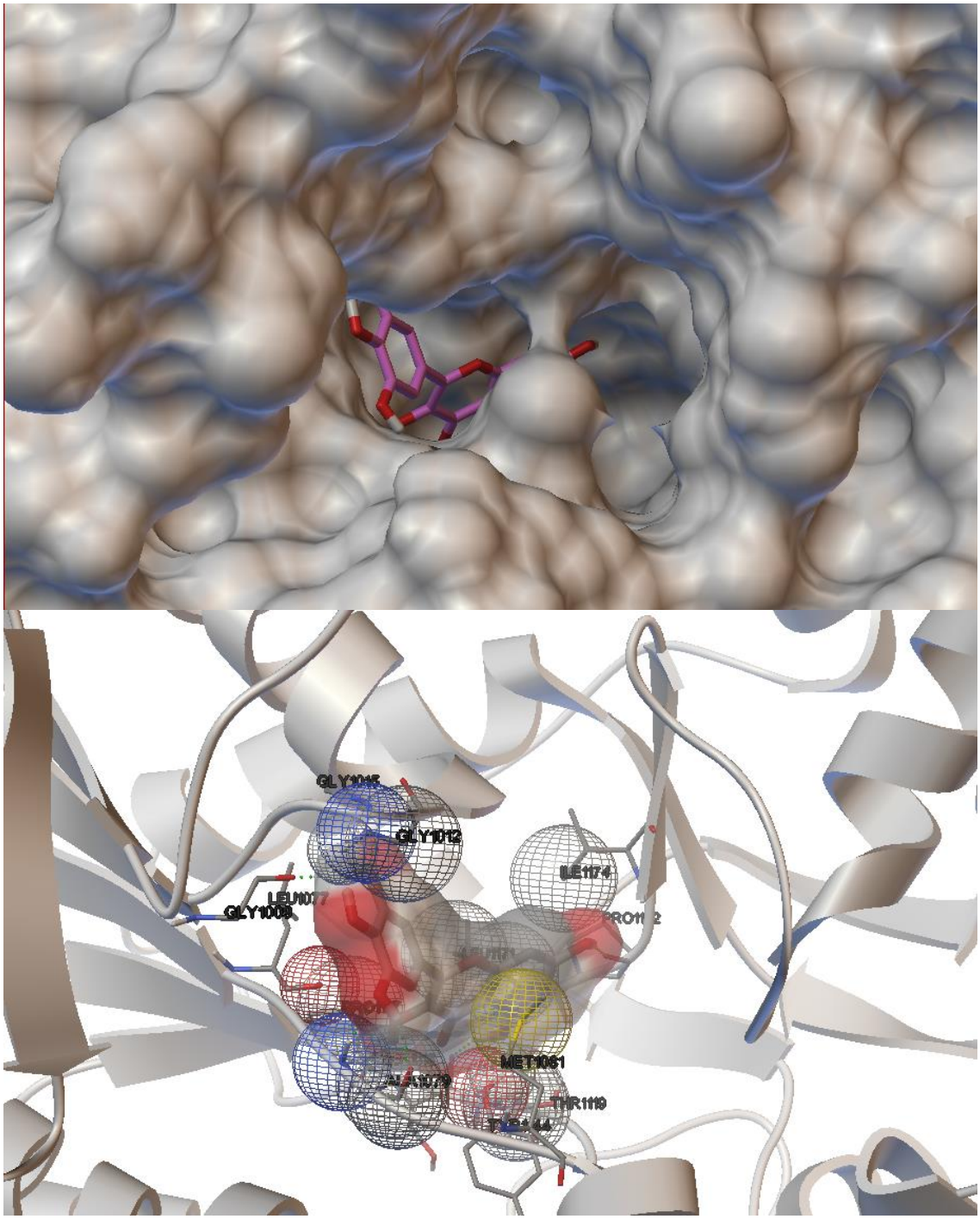


Figure 3.4.50 - binding pose of myricetin predicted by AutoDock in Virtual Screen 3. Myricetin is predicted to bind in the NAD binding pocket. It also interacts with important L-threonine-binding residues Met81, Thr119 and Tyr144.

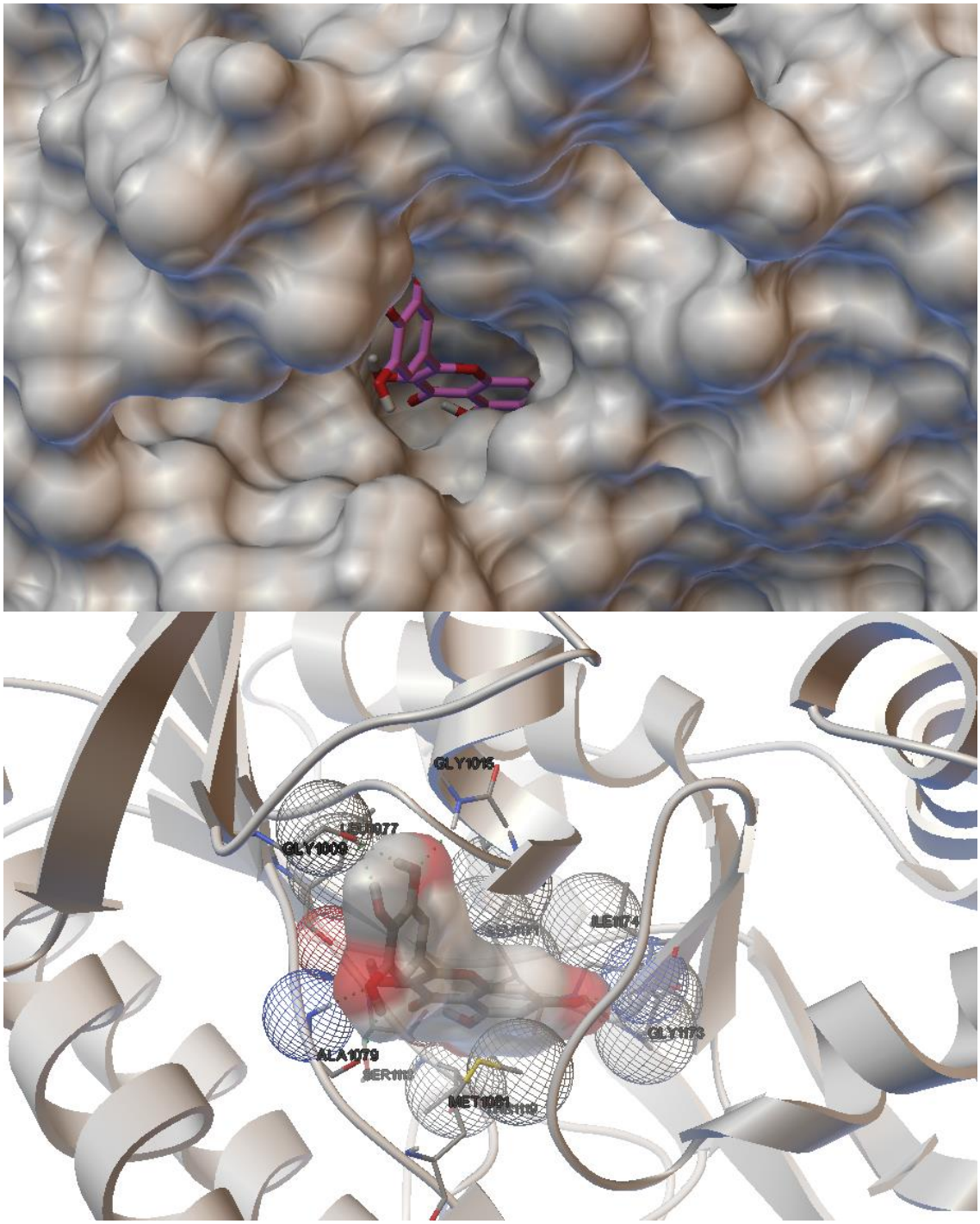


Figure 3.4.51 - binding pose of myricetin predicted by AutoDock in Virtual Screen 4. As in Virtual Screen 3, myricetin is predicted to bind in the NAD binding pocket. It also interacts with important L-threonine-binding residues Met81, and Thr119.

In Virtual Screens 3 and 4, myricetin was predicted to bind in the NAD^+ binding pocket, which conflicts with the conclusions of the MOI studies. This inhibitor will be discussed further in a later section.

NSC132249

The inhibitor of UDP-GalE from *Trypanosoma brucei* (*Tb*GalE) NSC132249, was one of the most potent compounds in its series and inhibited TDH with an IC_{50} of 52.2µM. A Hill coefficient of 1.33 was calculated for the slope fit to the plot of fractional activity as a function of inhibitor concentration. This suggests a minor degree of cooperativity may be involved in its binding.

The MOI of NSC132249 in relation to NAD^+ was predicted to be competitive, with a K_i of 67.5µM. The competitive model was deemed by an AIC comparison to be more likely to have generated the data than a mixed model (which predicted an alpha of 5.38, also suggesting a competitive MOI).

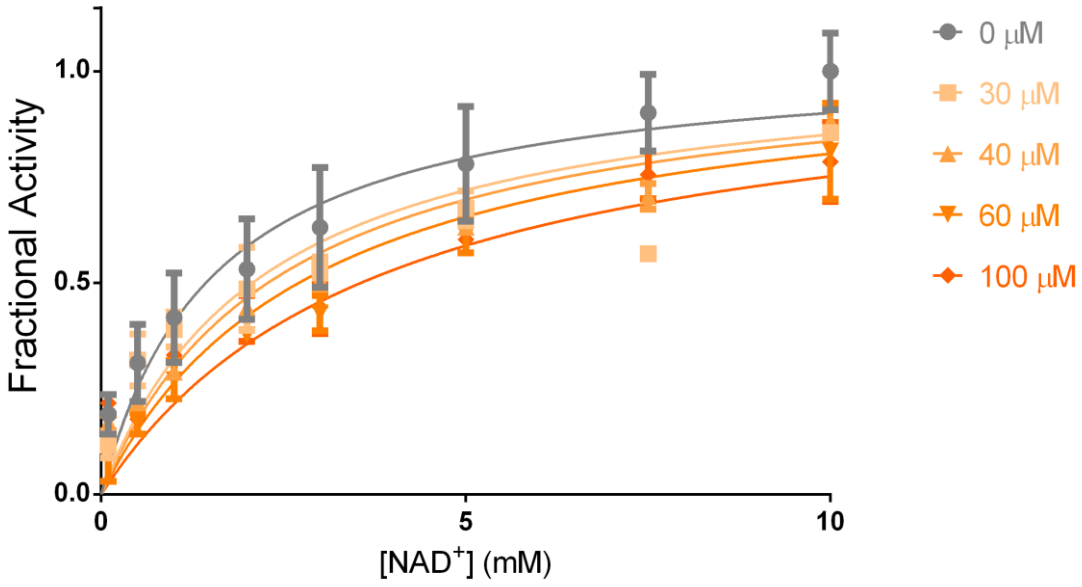


Figure 3.4.52 - curves describing competitive inhibition fitted to data describing the relationship between NAD^+ concentration and fractional activity, measured in the presence of different fixed concentrations of NSC132249.

The MOI of NSC132249 in relation to L-threonine was more ambiguous, with mixed, uncompetitive and noncompetitive models all providing similar fits to the data. Comparisons using AIC indicated that the uncompetitive model was the most appropriate. However, an F-test confirmed that the uncompetitive model was the most appropriate, but only with a P-value of 0.72. For this reason, a mixed noncompetitive-uncompetitive model, where the value of alpha, and hence the affinity for the free enzyme, is low is assigned to NSC132249 in relation to L-threonine. The calculated K_i is high at 687µM, as would be expected for an uncompetitive inhibitor. The αK_i is calculated as 119.6µM. This is higher than that which would be calculated from the IC_{50}, using the Cheng-Prusoff equation for uncompetitive inhibitors: 41.2µM. Once again, this is probably due to the over-estimation of V_{max} when fitting equations to data that do not strictly conform to Michaelis-Menten kinetics.

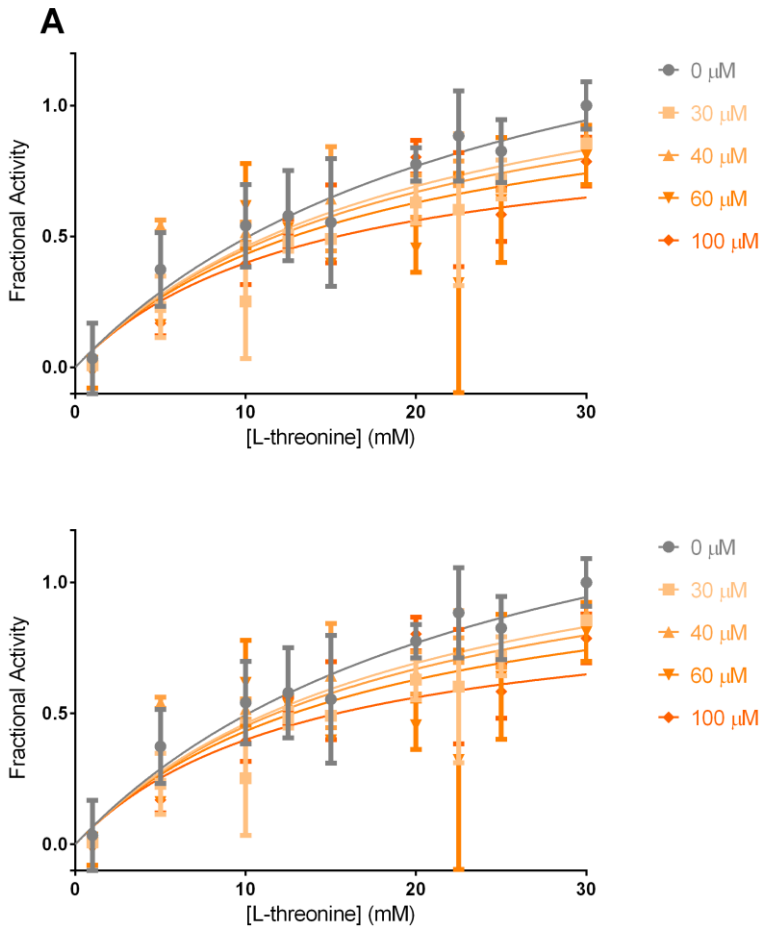


Figure 3.4.53 - plots of fractional activity against L-threonine concentration, at different fixed concentrations of NSC132249. Curves describing uncompetitive inhibition (A) and mixed inhibition (B).

The data suggest that NSC132249 inhibits TDH by binding to the NAD site and that this binding is more likely to occur when L-threonine is bound to the enzyme. Binding to the NAD binding site is supported by the predictions of AutoDock in Virtual Screen 4.

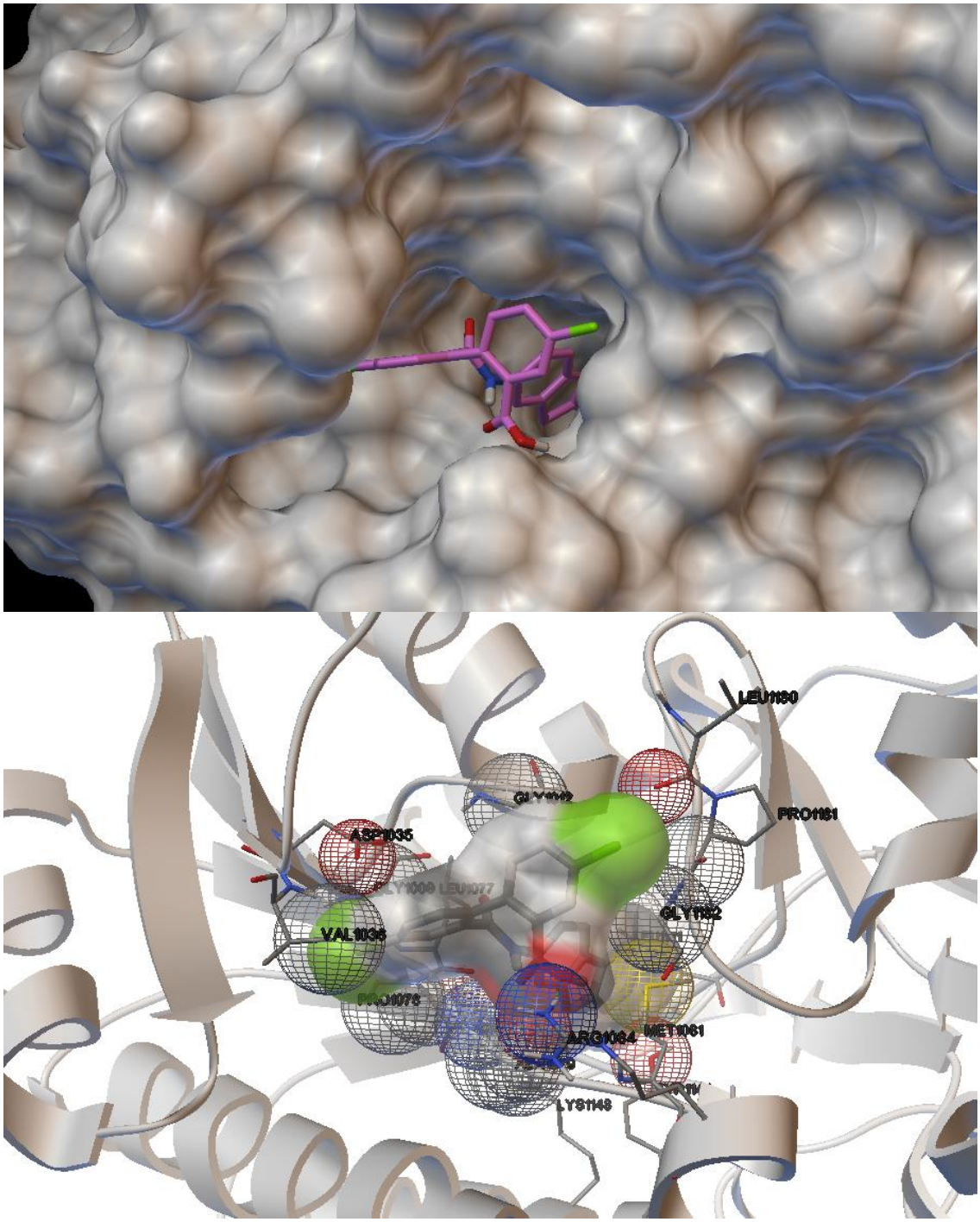


Figure 3.4.54 - binding mode of NSC132249 predicted by AutoDock in Virtual Screen 4. The compound occupies the NAD^+ binding site, but also interacts with Leu180, Pro181 and Gly182 of flexible loop 1, which is in its closed conformation.

An uncompetitive MOI in relation to L-threonine becomes more plausible when it is considered that the binding mode depicted in Figure 3.4.54 occurs when flexible loop 1 (TDH residues Thr179-Ala185) is in its closed conformation. The structural data presented herein suggest that this conformation may be promoted by L-threonine binding.

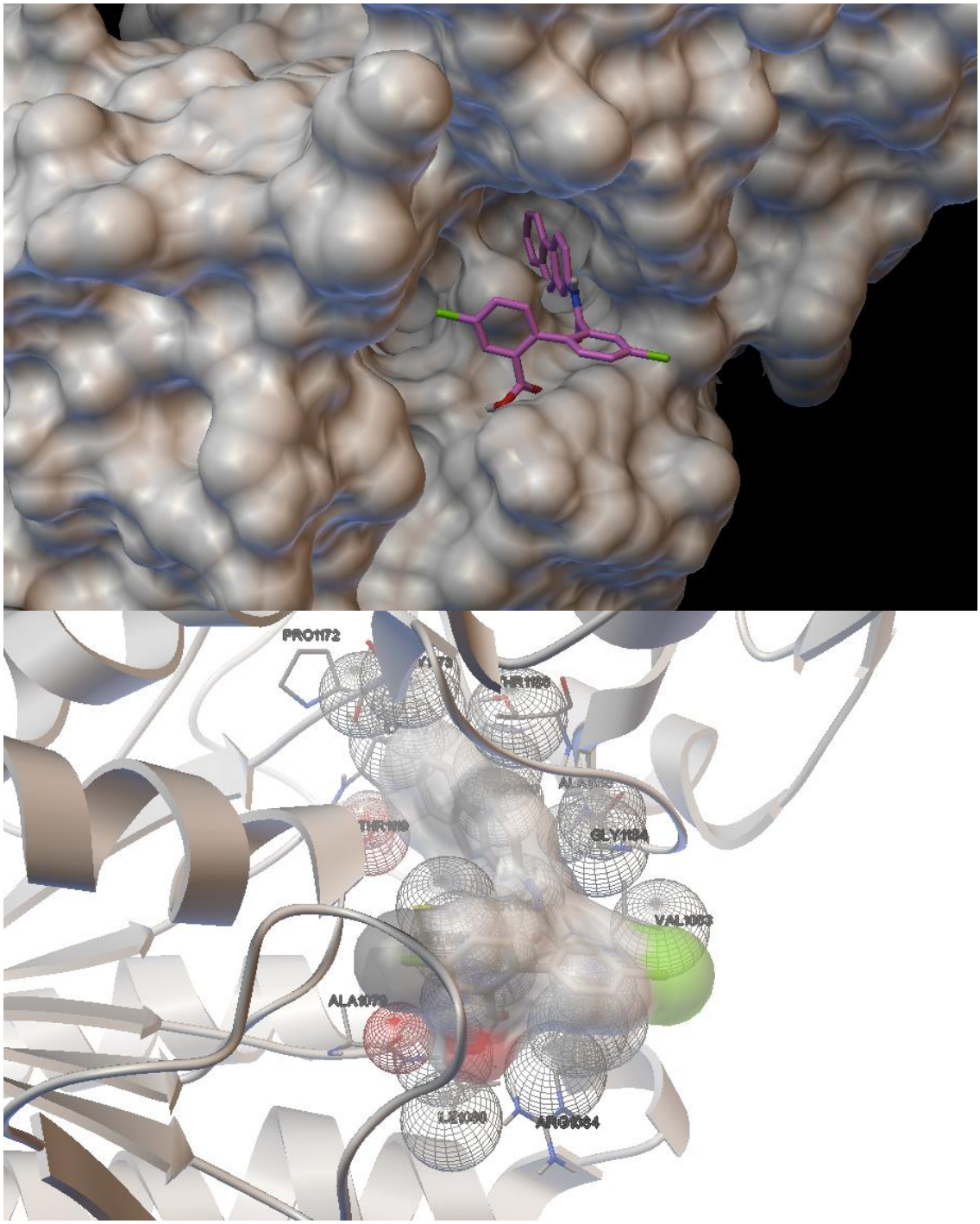


Figure 3.4.55 - the binding mode of NSC132249 predicted by AutoDock in Virtual Screen 3. NSC132249 is predicted to bind the L-threonine binding pocket. This conflicts with the prediction made in Virtual Screen 4, in which NSC132249 was also a hit.

The binding mode predicted in Virtual Screen 3 (see Figure 3.4.55, above) would appear to contradict the prediction of an uncompetitive MOI towards L-threonine because NSC132249 binds in the L-threonine pocket. Alternatively, taking into account the cooperative nature of the relationship between L-threonine concentration and TDH rate, it may be possible that binding of

L-threonine to the active site of one monomer, increases the affinity of the second monomer for NSC132249.

NSC132252

As discussed earlier, NSC132252 required further study because of its different behaviour in the presence or absence of the detergent Triton X-100. In contrast to the other compounds in the *Tb*GalE inhibitor series, the onset of its inhibition is rapid. However, it becomes a time-dependent inhibitor in the presence of Triton X-100. This suggests that it may act as a non-specific, promiscuous inhibitor unless this mechanism of inhibition is disrupted by detergent. Alternatively, the compound may possess two specific mechanisms of inhibition, one of which is disrupted by detergent. In the absence of detergent, the compound's IC_{50} is very low at 4.2µM. Conversely, the IC_{50} is somewhat higher in the presence of detergent: 21.0µM

NSC132252 in the absence of detergent

Titration of NSC132252 concentration produced data that fitted to a curve with a Hill coefficient of 1.7, suggesting that there may be some cooperativity involved in NSC13252 inhibition of TDH. The MOI of NSC132252 in relation to NAD^+ was determined to be competitive.

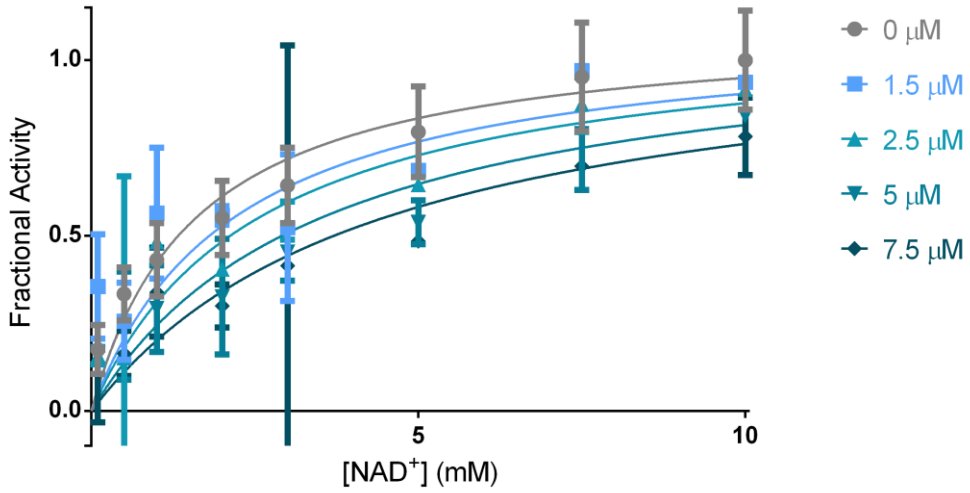


Figure 3.4.56 - plot of fractional activity versus NAD^+, at different fixed concentrations of NSC132252. Curves describing competitive inhibition are fit to the data.

Models for mixed, competitive and noncompetitive MOIs all fit the data relating to L-threonine concentration and rate with similar R^2 values of 0.7623 – 0.7626. The mixed model fit with an alpha value of 17.49, but with large margins of error. This suggests that the relevant MOI is competitive or mixed noncompetitive-competitive.

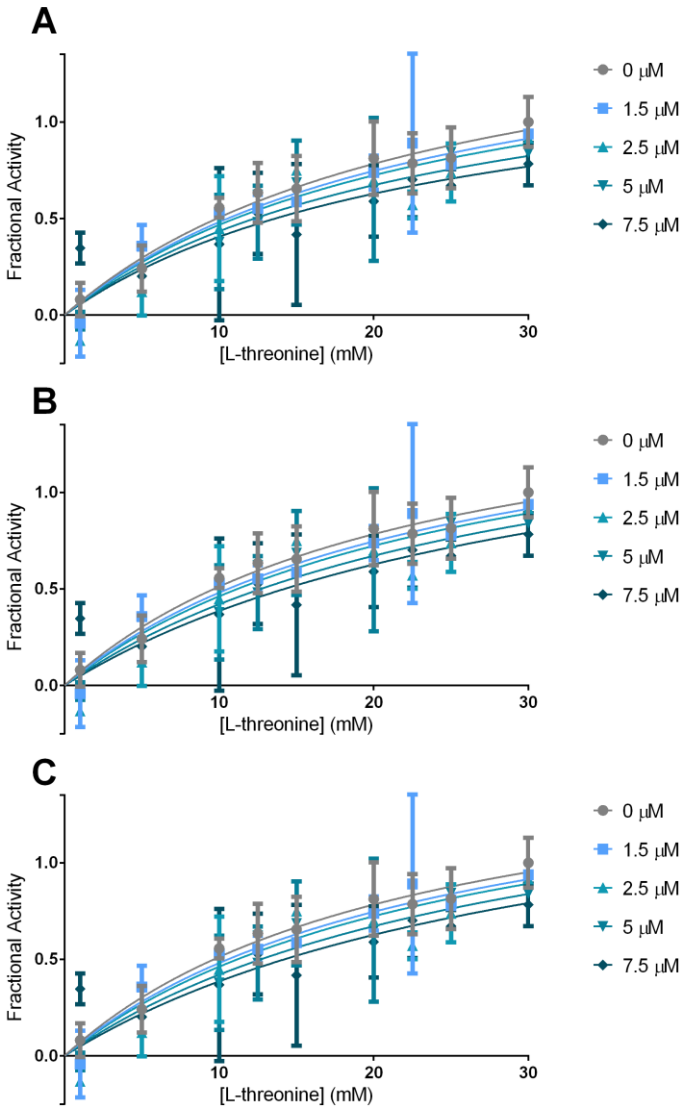


Figure 3.4.57 - plots of fractional activity as a function of L-threonine concentration, at different fixed concentrations of NSC132252. Curves based on mixed inhibition (A), competitive inhibition (B) and noncompetitive inhibition (C) all had similar fits to the data.

An AIC analysis compared fits of competitive and noncompetitive models and reported that the competitive model was more likely to have produced the data, but only with a relative probability of 1.11. Therefore, the MOI of NSC132252 in the absence of detergent is best assigned as a mixed noncompetitive-competitive MOI, where the inhibitor has a strong preference for binding when L-threonine is absent.

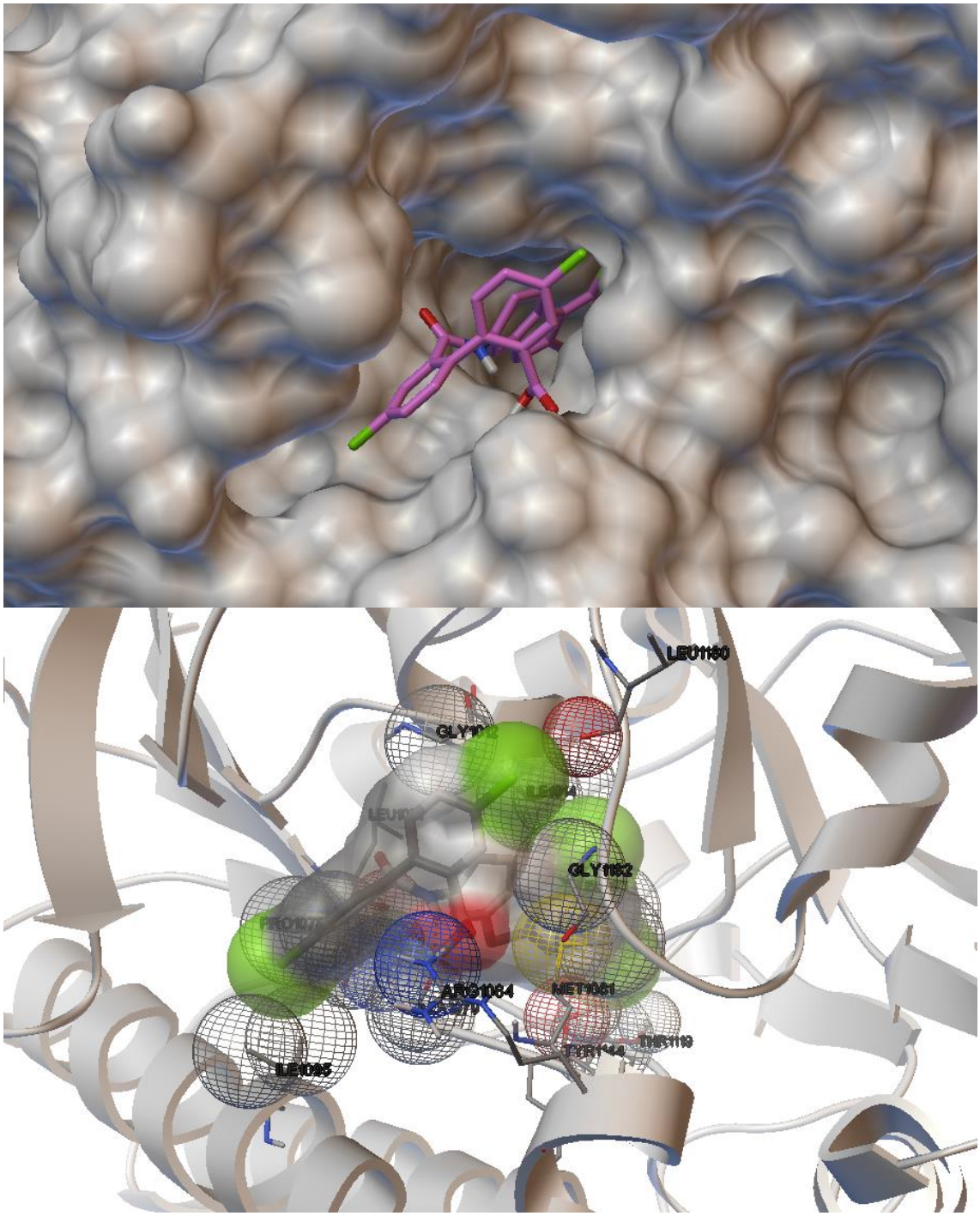


Figure 3.4.58 - predicted binding mode of NSC132252 in Virtual Screen 4. The compound is predicted to bind in both substrate binding sites, simultaneously

The data obtained do not suggest that the rapid-onset inhibition of TDH by NSC132252 is caused by a non-specific mechanism. One would expect a noncompetitive MOI if this were the case. Rather, the data indicate that NSC132252 may bind to both NAD^+ and L-threonine binding pockets, as predicted in Virtual Screen 4 (see Figure 3.4.58).

NSC132252 in the presence of detergent

In the presence of 0.01% Triton X-100, NSC132252 exhibited a time-dependent inhibition of TDH that was less potent than the rapid inhibition exhibited without detergent. A fit of the mixed model of inhibition to the data relating to NAD^+ resulted in an R^2 of 0.842. The alpha value was not much greater than 1, at 2.693, indicating the possibility that the MOI was noncompetitive or mixed noncompetitive-competitive towards NAD^+. Comparison of all possible models using AIC confirmed that noncompetitive is the most appropriate MOI to be assigned to NSC13252 in relation to NAD^+ in the presence of detergent. Hill-modified versions of the relevant equations brought statistically significant improvements to the fits. A Hill coefficient of 2.2 was calculated by fitting a Hill-modified noncompetitive model to the data (R^2 = 0.850). A K_i of 14.0µM was calculated.

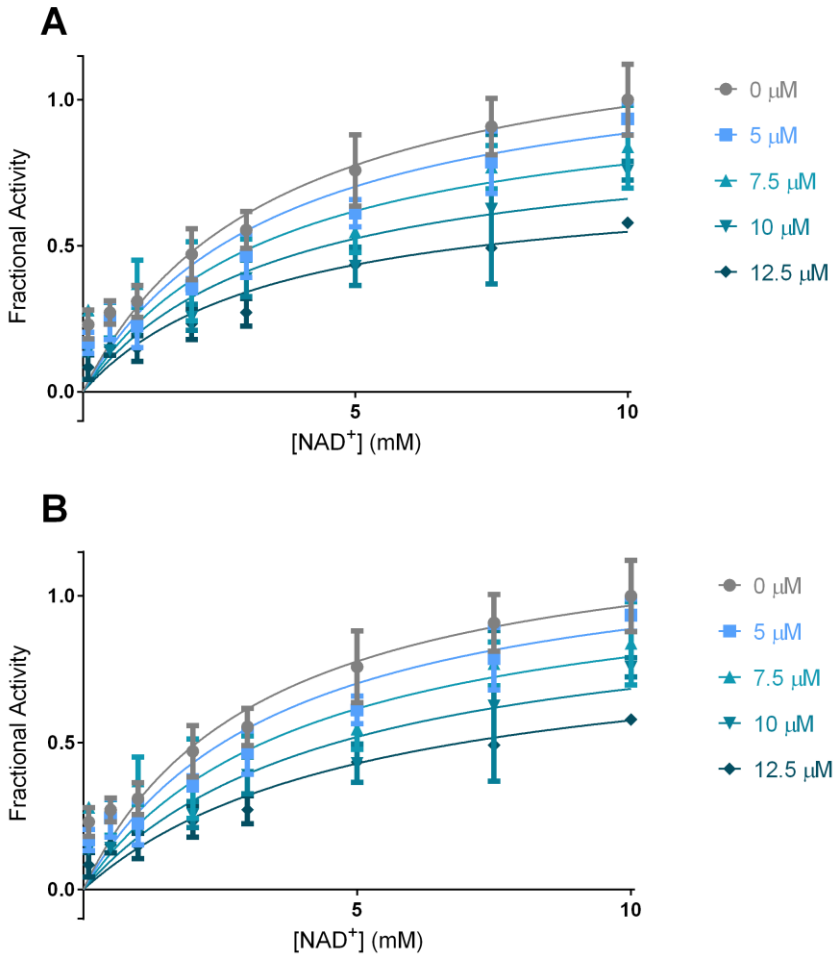


Figure 3.4.59 - plots of fractional activity against NAD^+ concentration, at fixed concentrations of NSC132252 and in the presence of 0.01% Triton X-100. Curves describing cooperative mixed inhibition (A) and cooperative noncompetitive inhibition (B) provided similar fits to the data.

The MOI for NSC132252 in the presence of detergent was determined to be competitive towards L-threonine. The mixed model of inhibition was used to calculate an alpha of 8.17, which indicates preferential binding to free enzyme. An AIC analysis also confirmed the competitive model to be the most accurate description of the MOI. In this case, the use of Hill-modified equations for cooperative inhibition did not provide any significant improvement to the fits of each model to the data.

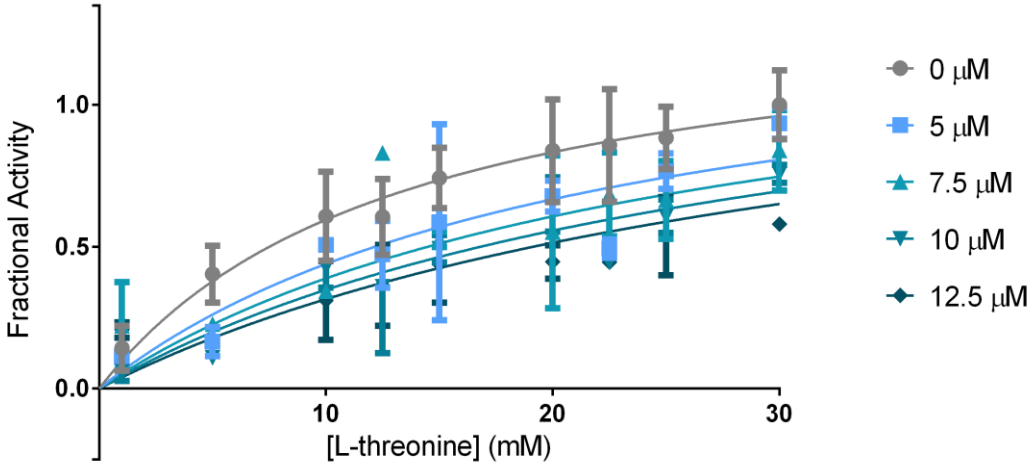


Figure 3.4.60 - fit of fractional activity versus L-threonine concentration, at different fixed concentrations of NSC132252 and in the presence of 0.01% Triton X-100

The data obtained in the presence of detergent suggest that NSC132252 acts by a MOI that primarily involves binding to TDH in a way that is mutually exclusive with L-threonine binding. However, the binding in relation to NAD^+ was judged to be noncompetitive, which contradicts the binding poses predicted in Virtual Screens 3 and 4, which predict binding to the NAD^+ binding site.

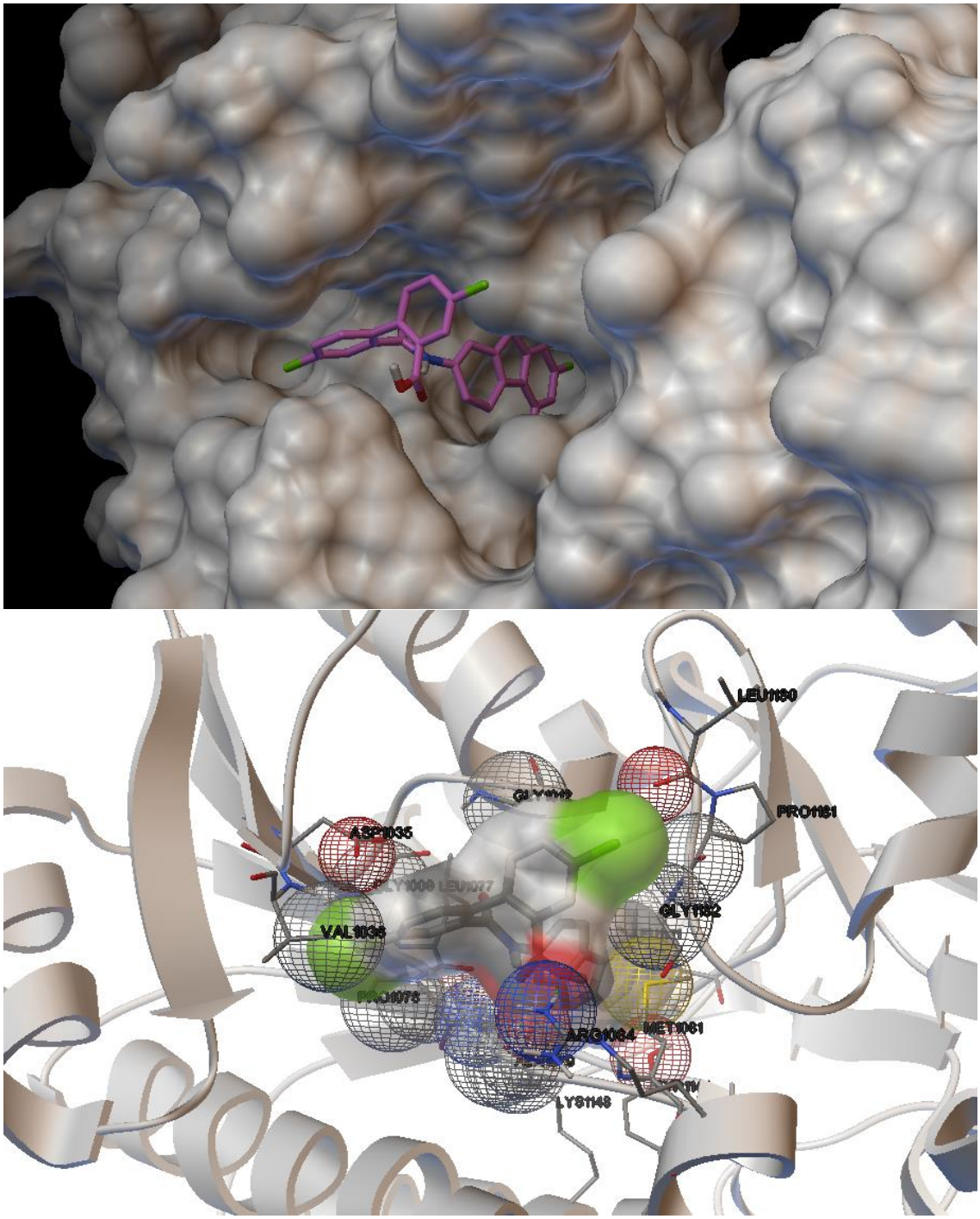


Figure 3.4.61 - predicted binding mode of NSC132252 in Virtual Screen 3. It is predicted to bind in the NAD^+ binding site. See also the predicted binding mode in Virtual Screen 4 (Figure 3.4.58).

The predicted MOIs of NSC132252 in the absence and presence of detergent are in part agreement with each other, in regards to L-threonine binding. However, the non-linear regression analyses predicted that NSC132252 may compete with NAD^+ in the absence of detergent, whereas NSC132252 appeared to be noncompetitive towards NAD^+ in the presence of detergent. There are a number of scenarios that can explain this difference. Firstly, the possibility that the MOIs towards both L-threonine and NAD^+ in the absence of detergent were

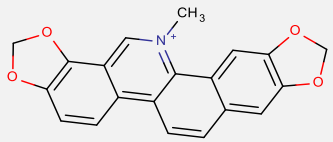
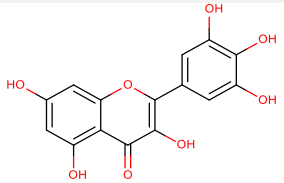
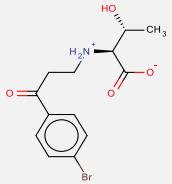
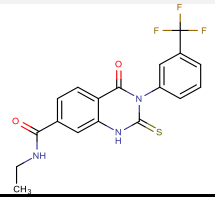
mixed noncompetitive-competitive, suggests that the mechanism of inhibition may not actually involve binding to TDH, but may impede binding of the substrates to the enzyme. Another scenario that was mentioned earlier is that the inhibition shown in the absence of detergent may actually be a combination of non-specific inhibition (i.e. that caused by aggregates) and the specific inhibition that is demonstrated in the presence of detergent. A third scenario, also alluded to earlier, is that NSC132252, may be able to bind TDH in two distinct binding modes. One of these binding modes could be disrupted by detergent, so the MOI of NSC132252 presents differently under different conditions.

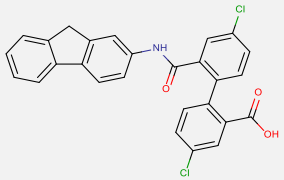
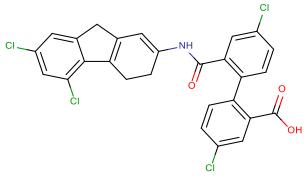
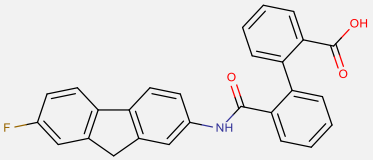
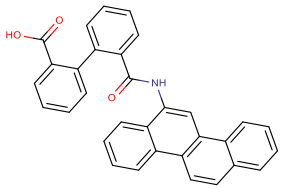
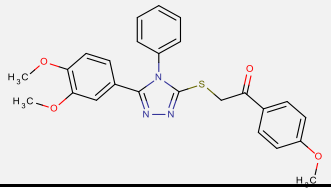
Compared to the other *Tb*GalE inhibitor, NSC132249, the MOI of NSC132252 is not entirely consistent. The data indicate that NSC132249 binds to the NAD^+ site, preferentially to L-threonine-bound TDH. The data on NSC132252 inhibition suggested that it may bind to the NAD^+ site in the absence of detergent. However, a noncompetitive MOI is implicated in the presence of detergent. The possibility of NSC132249 also binding to the L-threonine site was raised above. However, until crystallographic data can be gained on the actual binding positions of these inhibitors, the data on the MOIs of the *Tb*GalE inhibitors remains inconclusive.

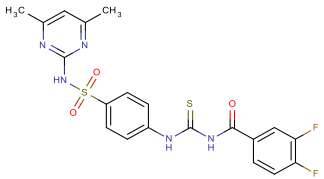
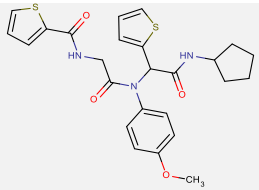
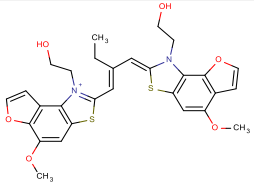
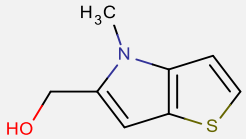
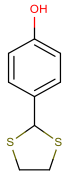
3.4.5 Inhibition of trypanosome growth

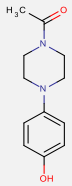
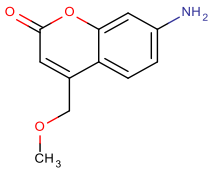
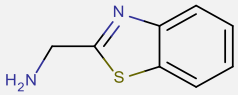
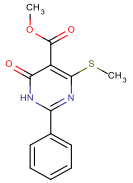
Hits from *in vitro* screening were tested for trypanosome growth inhibition by Professor Jon Kelly at the London School of Hygiene and Tropical Medicine, as described in Section 2.4.4. The results of a preliminary round of assays testing the *in vitro* screening hits for their ability to inhibit the growth of cultured bloodstream form (BSF) *T. brucei* are presented in Table 3.4.6, below.

Table 3.4.6 - the percentage inhibition of BSF trypanosome growth, in the presence of confirmed hits from the *in vitro* screening experiments. The names of the most potent compounds are emboldened and underlined.

				Percentage inhibition of trypanosome growth (at molar concentration)		
Compound Library	**ID**	**Structure**	**MW**	**10µg.ml^{-1}**	**5µg.ml^{-1}**	**1µg.ml^{-1}**
Custom Library	**sanguinarine**		394.3	100% (at 25.4µM)	100% (at 12.7µM)	100% (at 2.5µM)
	myricetin		318.2	100% (at 31.4µM)	100% (at 15.7µM)	16% (at 3.1µM)
	2-{[3-(4-bromophenyl)-3-oxopropyl]amino}-3-hydroxybutanoic acid (BPOB)		330.2	28% (at 30.2µM)	9% (at 15.1µM)	0% (at 3.0µM)
	Qc1		393.4	97% (at 25.4µM)	96% (at 12.7µM)	9% (at 2.5µM)

				Percentage inhibition of trypanosome growth (at molar concentration)		
Compound Library	**ID**	**Structure**	**MW**	**10µg.ml^{-1}**	**5µg.ml^{-1}**	**1µg.ml^{-1}**
	NSC 132249		474.3	79% (at 21.1µM)	25% (at 10.5µM)	0% (at 2.1µM)
	NSC 132252		545.2	52% (at 18.3µM)	9% (at 9.2µM)	0% (at 1.8µM)
	NSC 128598		423.4	72% (at 23.6µM)	11% (at 11.8µM)	0% (at 2.4µM)
	NSC 128608		467.5	65% (at 21.4µM)	28% (at 10.7µM)	0% (at 2.1µM)
Asinex	BAS 0792599		461.5	34% (at 21.7µM)	0% (at 10.8µM)	0% (at 2.2µM)

				Percentage inhibition of trypanosome growth (at molar concentration)		
Compound Library	**ID**	**Structure**	**MW**	**10µg.ml^{-1}**	**5µg.ml^{-1}**	**1µg.ml^{-1}**
	BAS 1066901		477.5	0% (at 20.9µM)	0% (at 10.5µM)	0% (at 2.1µM)
	ASN 5580922		497.6	32% (at 20.1µM)	0% (at 10.0µM)	0% (at 2.0µM)
	BAS 0472148		565.7	n.d.	98% (at 8.8µM)	0% (at 4.4µM)
Maybridge	CC 40909		167.2	24% (at 59.8µM)	0% (at 29.9µM)	0% (at 6.0µM)
	SPB 00826		198.3	51% (at 50.4µM)	10% (at 25.2µM)	0% (at 5.0µM)

				Percentage inhibition of trypanosome growth (at molar concentration)		
Compound Library	**ID**	**Structure**	**MW**	**10µg.ml^{-1}**	**5µg.ml^{-1}**	**1µg.ml^{-1}**
	SB 02047		220.3	60% (at 45.4µM)	0% (at 22.7µM)	0% (at 4.5µM)
	RJF 01106		205.2	42% (at 48.7µM)	0% (at 20.4µM)	0% (at 4.9µM)
	CC 06013	HCl 	200.7	26% (at 49.8µM)	0% (at 24.9µM)	0% (at 5.0µM)
	BTB 08846		276.3	51% (at 31.2µM)	9% (at 18.1µM)	0% (at 3.6µM)

The compounds with the most potent inhibitory activity were sanguinarine, myricetin and Qc1. Sanguinarine was particularly potent and inhibited trypanosome growth by 100% at 1µg/ml. Only one of the Asinex compounds exhibited significant inhibition of trypanosome growth – BAS 0472148 caused 98% inhibition at 5µg/ml, which suggests that it may possess potency that is comparable with myricetin and Qc1. None of the fragments, which would be expected to have lower potency at the drug target level, showed significant inhibition below 10µg/ml. BPOB appears to be a weak inhibitor of BSF trypanosome growth. Meanwhile, the *Tb*GalE inhibitors, NSC132249, NSC132252, NSC128598 and NSC128608, demonstrated comparable inhibition, and inhibited BSF trypanosome growth modestly at 5µg/ml. The compounds sanguinarine, myricetin and Qc1 were tested at a range of concentrations and in triplicate to more accurately determine the potency of BSF *T. brucei* inhibition. The IC_{50}s and concentrations causing 90% inhibition (IC_{90}) are displayed in Table 3.4.7.

Table 3.4.7 - BSF T. brucei growth inhibitory concentrations of sanguinarine, myricetin and Qc1.

Compound	IC_{50} ± SD (µM)	IC_{90} ± SD (µM)
sanguinarine	0.41 ± 0.04	0.48 ± 0.02
myricetin	10.90 ± 0.20	12.30 ± 0.20
Qc1	7.80 ± 0.36	11.1 ± 0.7

In the second trypanosome growth inhibition assay, the potency of sanguinarine was demonstrated to be very high, and significantly lower than that detected in the preliminary assay. The potencies demonstrated by myricetin and Qc1 were consistent with those observed in the preliminary assay. As some of the most potent inhibitors of TDH, the antitrypanosomal activity of sanguinarine, myricetin, Qc1 and BAS 0472148 is an encouraging sign of TDH-mediated efficacy. However, their relative degrees of trypanosome growth inhibition do not correlate entirely with their relative TDH inhibitory potencies. These findings will be discussed in more depth in section 4.

4. Discussion

4.1 TDH structure and function

4.1.1 *Tb*TDH as a GalE-like TDH

L-threonine 3-dehydrogenase from *Trypanosoma brucei* (TDH or *Tb*TDH) is a UDP-galactose 4'-epimerase-like (GalE-like) TDH, sharing many features with the enzymes characterised from *Flavobacterium frigidimaris* (*Ff*TDH), *Cupriavidus necator* (*Cn*TDH) and *Thermoplasma volcanium* (*Tv*TDH) in recent years(124–126,197). Three-dimensional structures of TDH determined using X-ray crystallography revealed that *Tb*TDH possesses features that are conserved amongst these enzymes, as well as other enzymes within the short-chain dehydrogenase/reductase (SDR) enzyme superfamily: a monomeric length of around 320 residues, a glycine-rich region (GxxGxxG) within the NAD-binding domain, a structural motif including lysine and tyrosine (YxxxK) near the active site, and several other conserved residues in the active site (see Sections 3.1.3 and 3.1.5). The TDH studied in this thesis appears to differ significantly from TDH in other species, such as *E. coli*, *Pyrococcus horikoshii* and *Thermococcus kodakarensis*, which are homotetrameric enzymes, rather than homodimeric, and which specifically bind divalent metal cations such as zinc(104,121,282). The GalE-like TDHs are distinguished from GalE by the presence of a loop adjacent to the active site – this is the flexible Loop 1 in *Tb*TDH. At the end of this loop, a consensus sequence that may be used to further distinguish GalE-like TDH from GalE is a motif represented by the sequence GTTDY(125). In the trypanosomal TDH enzymes, this sequence is replaced by GADDY, but it is unclear if this confers any difference between the enzymes from those organisms and the other known GalE-like TDHs. The structural characteristics place TDH in the 'extended' category of SDRs (eSDR), and thus the enzyme is assigned the identifier SDR14E under the SDR nomenclature initiative(283).

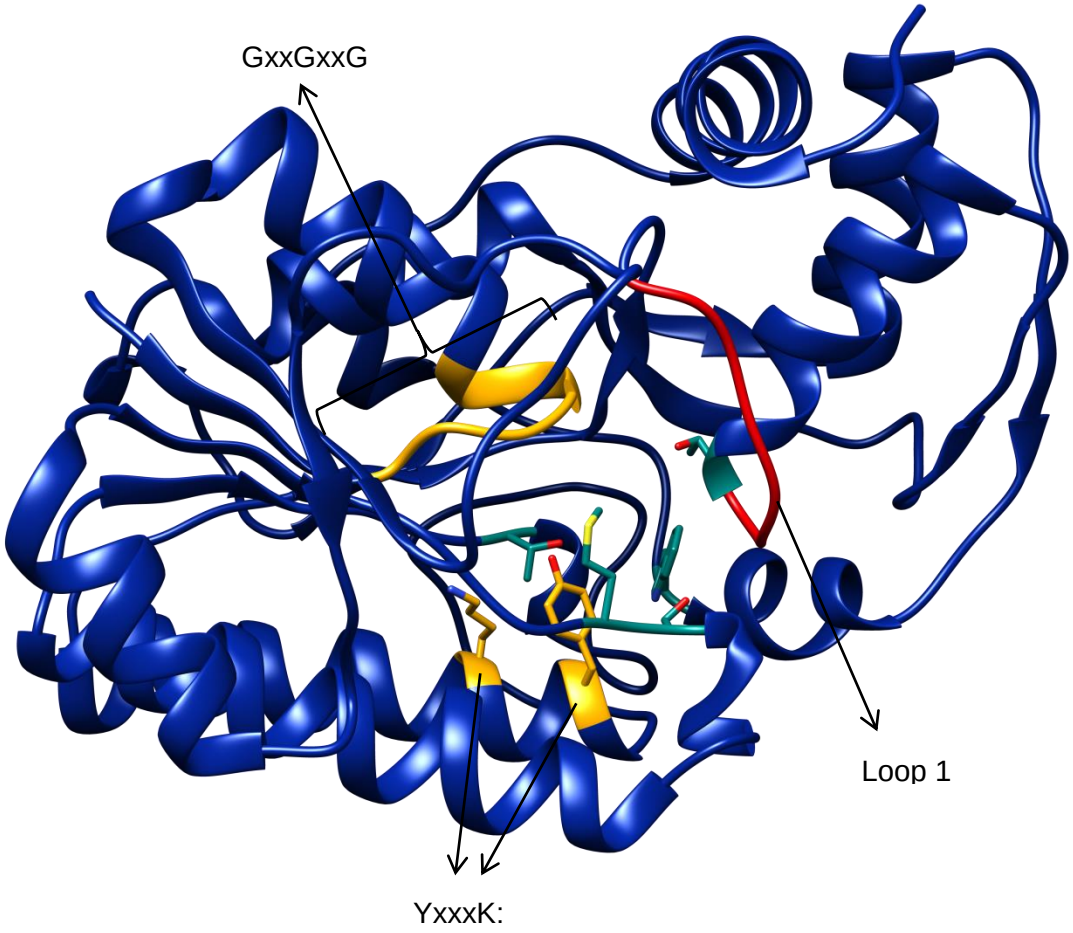


Figure 4.1.1 - *Tb*TDH monomer with the conserved features of GalE-like TDHs highlighted by colouring. The conserved sequences GxxGxxG and YxxxK are coloured gold, and other conserved active site residues are coloured cyan. The flexible loop 1 is coloured red.

The TDH enzyme studied in this thesis was shown to be a homodimer, and the interface between each TDH subunit was shown to be stabilised by a number of Van der Waals interactions and hydrogen bonds (see Section 3.1.6). A comparison of data from several X-ray crystallography experiments indicated that there may be some conformational flexibility in the regions making up the dimerisation interface. Therefore, it is likely that hydrogen bonds and salt bridges between the subunits are important in stabilising the dimer. Indeed, the same has been hypothesised about other GalE-like TDHs in the literature. Yoneda and colleagues demonstrated how the presence of a greater number of inter-subunit ion pairs in the dimers of *Tv*TDH compared to *Ff*TDH accounted for the greater thermostability of the former enzyme.

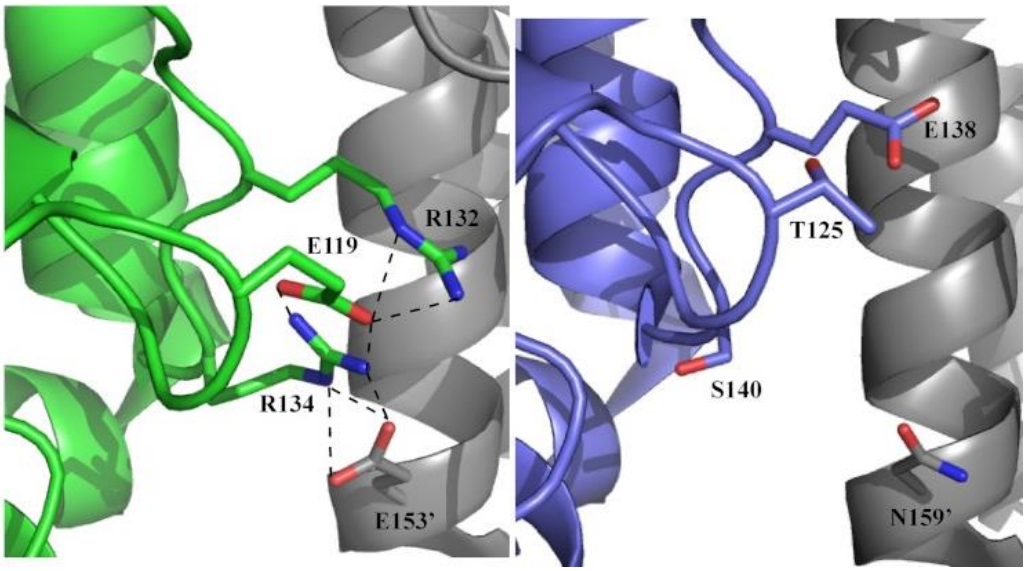


Figure 4.1.2 - comparison of the dimerisation interfaces of *Tv*TDH (left) and *Ff*TDH (right). The original images were presented by Yoneda and colleagues(125).

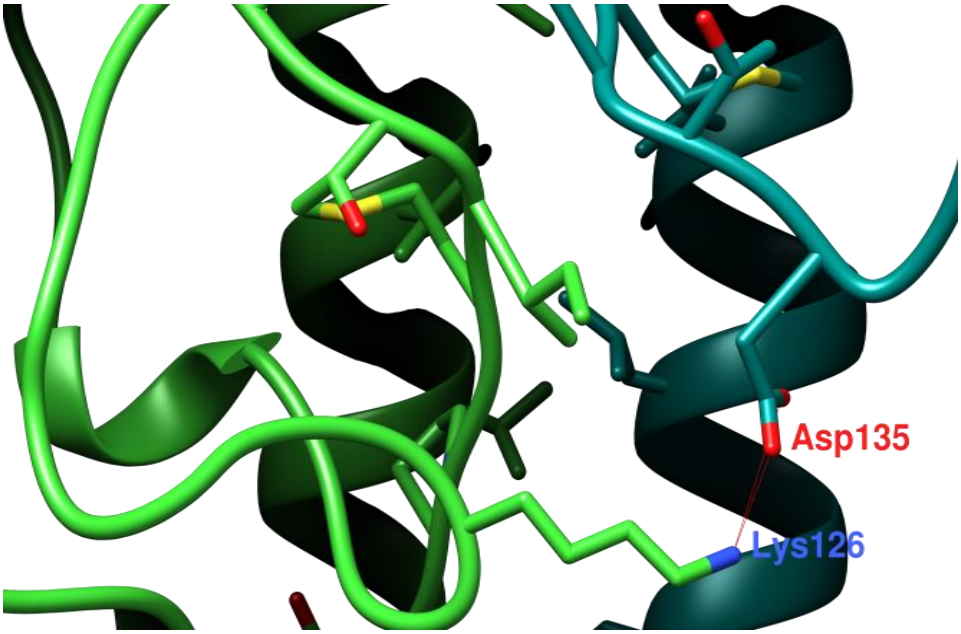


Figure 4.1.3 - the dimerisation interface of *Tb*TDH, as determined by X-ray crystallography. The image focuses on the equivalent location to those displayed in Figure 4.1.2.

Looking at Figure 4.1.2 and Figure 4.1.3, a comparison of the dimer interface of *Tb*TDH and *Tv*TDH shows that different inter-subunit ionic pairs stabilise the dimers in each enzyme. For example, Glu115 in *Tv*TDH is replaced by Lys126 in *Tb*TDH. Arg132 and Arg134 in *Tv*TDH are replaced by an asparagine and serine residue, respectively, which do not appear to make intra-residue contacts. Thus, it is reasonable to suggest that *Tb*TDH's lack of thermal stability (compared to *Tv*TDH) can be attributed to weaker, or less numerous inter-subunit hydrogen/ionic bonds stabilising the dimer.

Loops analogous to Loop 1 (see Section 3.1.5) identified in *Tb*TDH have been highlighted in TDH from other species(124,284). Indeed, the amino acid sequence corresponding to the loop appears to be a conserved feature among GalE-like TDHs. The studies here provided crystallographic evidence of distinct conformations adopted by Loop 1. The difference between them is characterised by a dramatic shift of up to 8 Å in distance, and the two conformations are easily classified as 'open' and 'closed'. The data here show several combinations of TDH conformation and enzyme form:

- *Apo* TDH in an open conformation;
- NAD^+-bound TDH in an open conformation;
- NAD^+-bound TDH in a closed conformation;
- NADH- and L-threonine- bound TDH in a closed conformation;
- NAD^+- and L-*allo*-threonine-bound TDH in an open conformation.

Different combinations of TDH conformations have been seen in dimers:

- TDH dimer with both subunits in the open conformation;
- TDH dimer with one subunit in the open conformation and one subunit in the closed conformation;
- TDH dimer with both subunits in the closed conformation.

Although the data are not conclusive, one can hypothesise that Loop 1 can be in an open or closed conformation, regardless of whether L-threonine is bound or not. It is possible that a change to a closed conformation is induced by binding of a ligand to the L-threonine binding site, as was observed with L-threonine- and pyruvate-bound structures. However, the binding of L-allo-threonine did not necessarily induce a closed structure. In consideration of the conformations adopted in different subunits of the TDH dimer, the data suggests that either the Loop 1 conformation in each monomer is adopted independently of the conformation in the other subunit, or that there may be a mechanism whereby the conformation of one subunit effects the conformation in the other subunit. As will be discussed further below, it is possible that Loop 1 plays an important role in the mechanism of TDH binding to its ligands. The loop remains open before and after NAD^+ binding to accommodate L-threonine, but then closes over the active site once L-threonine is bound.

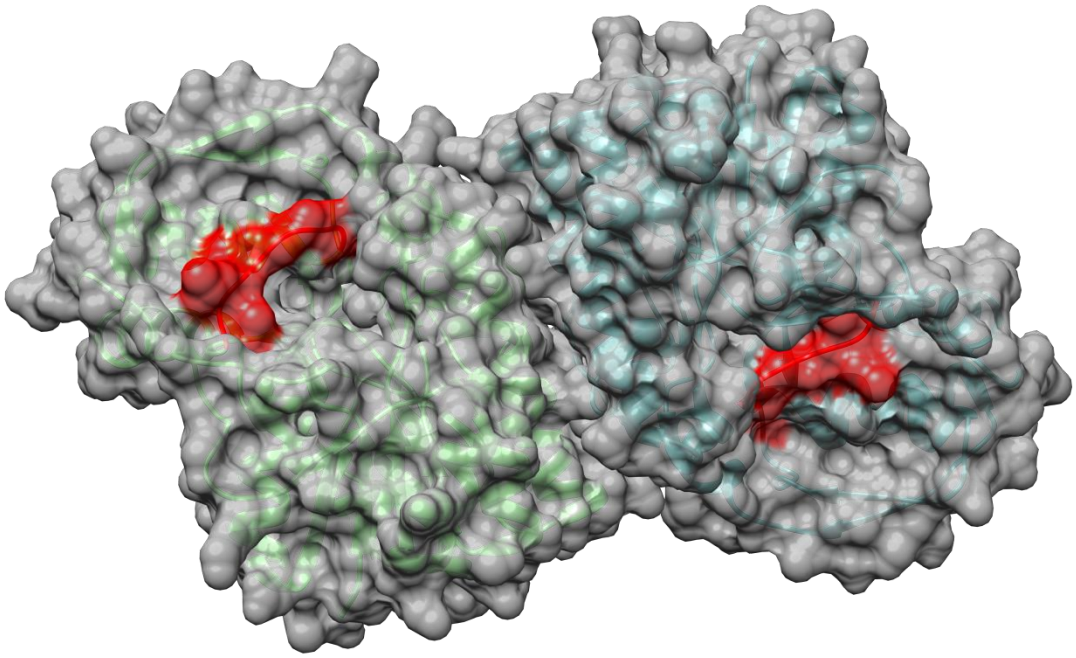


Figure 4.1.4 - combined ribbon and surface representation of the structure of the TDH dimer, determined in crystallographic model TQ5. In one subunit (green ribbon, left) Loop 1 is in the open conformation. In the second subunit (blue ribbon, right) Loop 1 is in the closed conformation. Loop 1 is coloured red in both subunits.

Recently, data from studies of TDH from *Cupriavidus necator* have also demonstrated that this loop region adopts different conformations in 'open' and 'closed' structures of TDH. Additionally, the authors of that study have suggested that the changes are induced by NAD^+ binding(284).

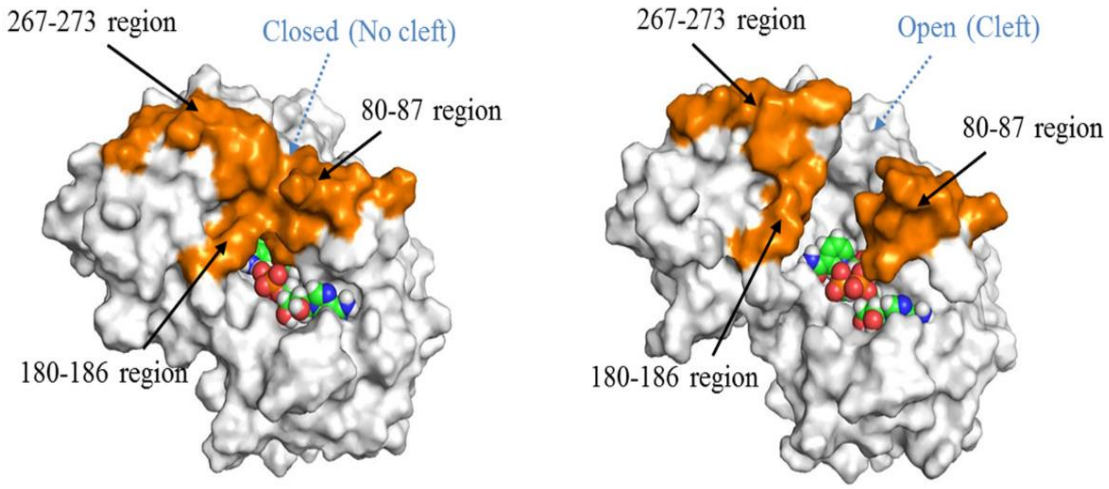


Figure 4.1.5 - surface representations of the structures of the *holo* form of *Cn*TDH at the start (left image) and end (right image) of a molecular dynamics simulation. Flexible regions are coloured orange. It can be seen that the flexible loop region at residues 180-186 is closed over the active site in the image on the left, whilst it is in an open conformation in the image on the right. Image taken from the study of Nakano et al.(284)

A combination of crystallographic data from *apo* and *holo* forms of *Cn*TDH, molecular dynamics simulations and enzyme kinetics were used to build a hypothesis of ligand binding and structural changes in this enzyme. Nakano et al. proposed a model whereby NAD^+ binding caused a rigidification of flexible regions in the NAD-binding domain, but Loop 1 remained open. Then, L-threonine binding causes Loop 1 to close over the active site, and this accounts for the selectivity of the enzyme(284).

A detailed examination of the conserved features in *Tb*TDH has also provided strong evidence that several other TDHs described in the literature are also GalE-like TDHs. TDH from chicken liver, pig liver and *Clostridium sticklandii* have all been described as homodimers with similar molecular weights to *Tb*TDH(259,260,285). In addition to these enzymes, TDH from *Capra hircus* (goat), *Mus musculus* and *Staphylococcus aureus* also appear to share features with *Tb*TDH and other GalEs. Findings related to TDH structure and function will therefore be compared with findings from studies of related TDH enzymes in other species.

4.1.2 The relationship between TDH structure and function

By successfully collecting multiple crystallographic data sets, including those of ligand-bound TDH, more insights into the enzyme's structure and function have been obtained. The TDH cofactor NAD, was bound in several of the TDH structure models described here. The nicotinamide ring was bound in a *syn* conformation, meaning that the hydride is transferred to the pro-S position on the C4 carbon (see Figure 3.1.37, Figure 3.1.38 and Figure 4.1.6). It has been proposed that, in alcohol dehydrogenases, the conformation in which a particular enzyme binds NAD (*anti* or *syn*) is the result of an evolutionary selection, whereby said conformation depends on the reactivity of the second ligand. Benner (1982) describes how a NAD conformation with weaker reducing power is used to reduce a relatively more reactive carbonyl. This supposedly helps to equalise the free energies of the bound substrates, thereby increasing catalytic efficiency. *Syn*-NADH is a stronger reducing agent than *anti*-NADH, so it is used to reduce more reactive carbonyls(286). Likewise, for TDH, the *syn* NAD, a weaker oxidising agent than *anti*-NAD, is balanced with a relatively less reactive carbonyl in 2-amino-3-ketobutyrate (AKB).

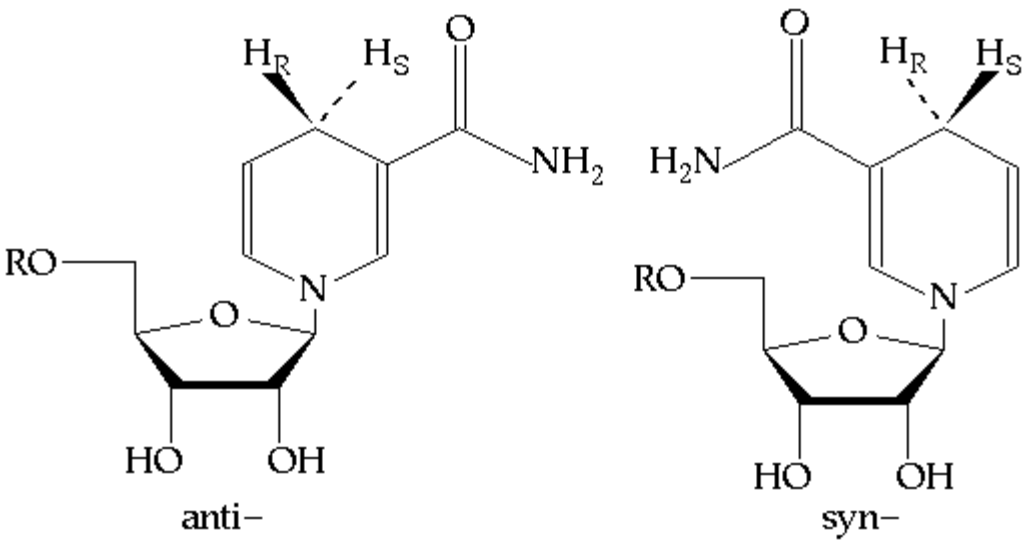


Figure 4.1.6 - *syn* and *anti* configurations of NAD(H). Image re-used from http://iop.vast.ac.vn/theor/conferences/smp/1st/kaminuma/UCSFComputerGraphicsLab/DH.html.

The NADH and L-threonine-bound structure presented here (see Figure 3.1.39 and Figure 4.1.8) is only accompanied by one other threonine-bound wild-type TDH structure(284). Both structures provide valuable insights into the mechanism of action of TDH. Principally, the proposed mechanism of action, in which a hydride is transferred from the β-carbon of L-threonine to the C4 atom of the nicotinamide ring of NAD appears to be confirmed, as all atoms lie in close proximity with each other.

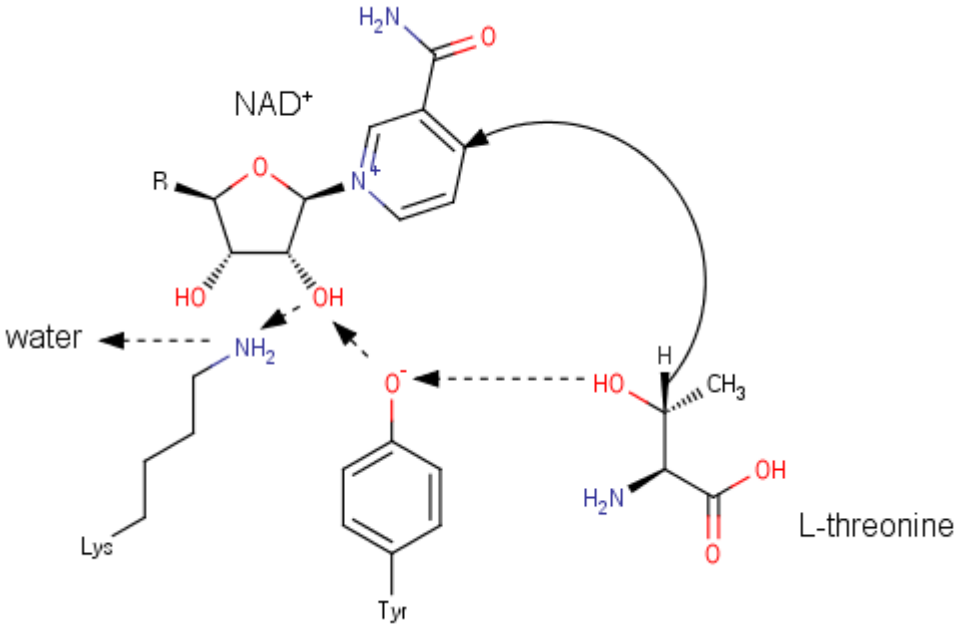


Figure 4.1.7 - schematic of the proposed mechanism of L-threonine oxidation by TDH. The curved arrow indicates transfer of a hydride from the ß-carbon of L-threonine to NAD. The dashed arrows indicate the proposed transfer of the proton from the ß-hydroxyl group of L-threonine to Tyr144, to NAD, to Lys148 and then onto a water molecule. This mechanism is supported by the crystallographic data.

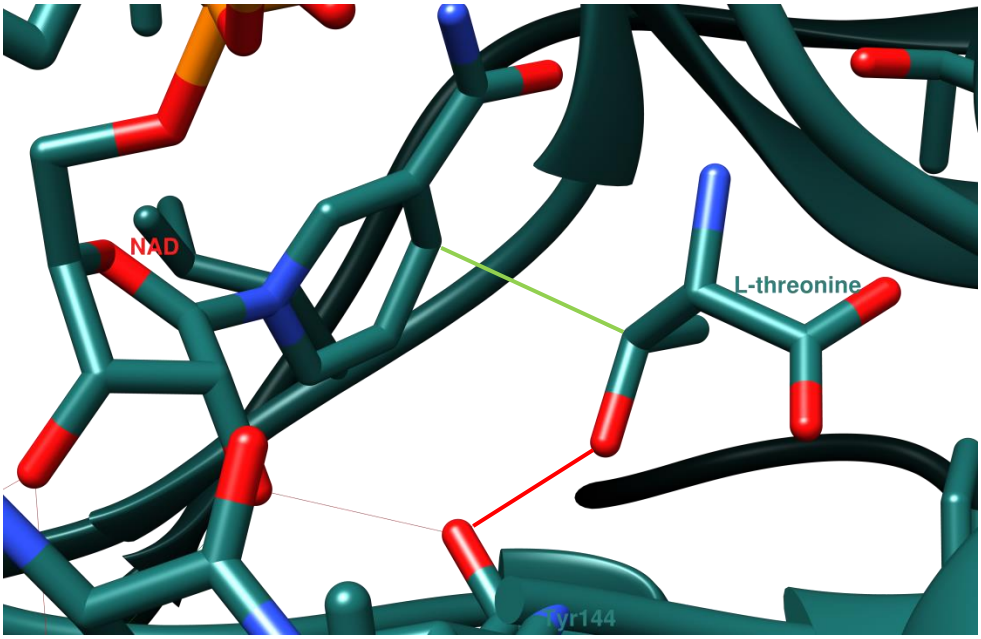


Figure 4.1.8 - close up of TDH structure in model TNADH4. L-threonine is bound in the active site and is situated close to the nicotinamide ring of NAD. The green line shows the path from the β-hydrogen of L-threonine to the C4 carbon of the nicotinamide ring. The red line highlights the path from the proton of the ß-hydroxyl group of L-threonine, to the phenyl group of Tyr144 and then to the hydroxyl group of the NAD ribose.

Hypotheses of roles played by important active site residues also gain support from the findings from crystal structure models of *Tb*TDH. Whilst residues such as Ser82 and Thr186 are shown to be binding partners for L-threonine, it is also plausible that Met81 and Trp280 play an important role in the correct positioning of L-threonine (see Figure 3.1.39). Perhaps more importantly, the residues of the "catalytic triad"(287,288), which have been described in alcohol dehydrogenases and SDRs in general, have been identified as Thr119, Tyr144 and Lys148 in TDH. The residues Thr119 and Tyr144 appear to hydrogen-bond the side chain hydroxyl group of L-threonine. The interatomic distances between the oxygen atoms of the relevant groups on L-threonine and Tyr144 also seem to support the hypothesis that Tyr144 acts as a base and withdraws a proton from the substrate. Molecular dynamics simulations in *Drosophila* alcohol dehydrogenase have demonstrated that this is made possible by the fact that the protonation states of the active site tyrosine and lysine residues are coupled(287–289). The side chains of both groups will be deprotonated at the same time, as depicted in Figure 4.1.7. Extrapolating this theory to the TDH mechanism of action, the lysine residue extracts a proton from a hydroxyl group on a NAD ribose group, which in turn extracts the proton from the tyrosine residue, thus allowing it to extract the proton from L-threonine. In alcohol dehydrogenase this proton is then proposed to be transferred to a chain of eight water molecules, which relay the proton until it is removed from the interior of the enzyme. In this way, the tyrosine residue is deprotonated again and is able to participate in the catalysis of another reaction(289,290). The interatomic distances between the relevant groups observed in TDH structure models all seem to lend support to this theory. The pH dependence of the protonation state of lysine and tyrosine may

also explain why the optimum pH in *Tb*TDH and other GalE-like TDHs is around pH 8-9(125,126,139,197,198,202,285,291). A short chain of water molecules close to Lys148 was identified in several of the TDH structures obtained here, but a complete chain of eight water molecules, which has been shown to be essential for ADH activity(290), was not observed. There are a number of explanations for this. Firstly, it may be that a shorter water chain is necessary for activity in TDH. Secondly, it may be that the occupancies of other water molecules in the chain were not high enough for their positions to be determined in any of the TDH structure models solved. Thirdly, a complete water chain may only be formed during catalysis, and the conformational state of TDH in which this water chain is complete has not been captured in any of the crystallographic models to date.

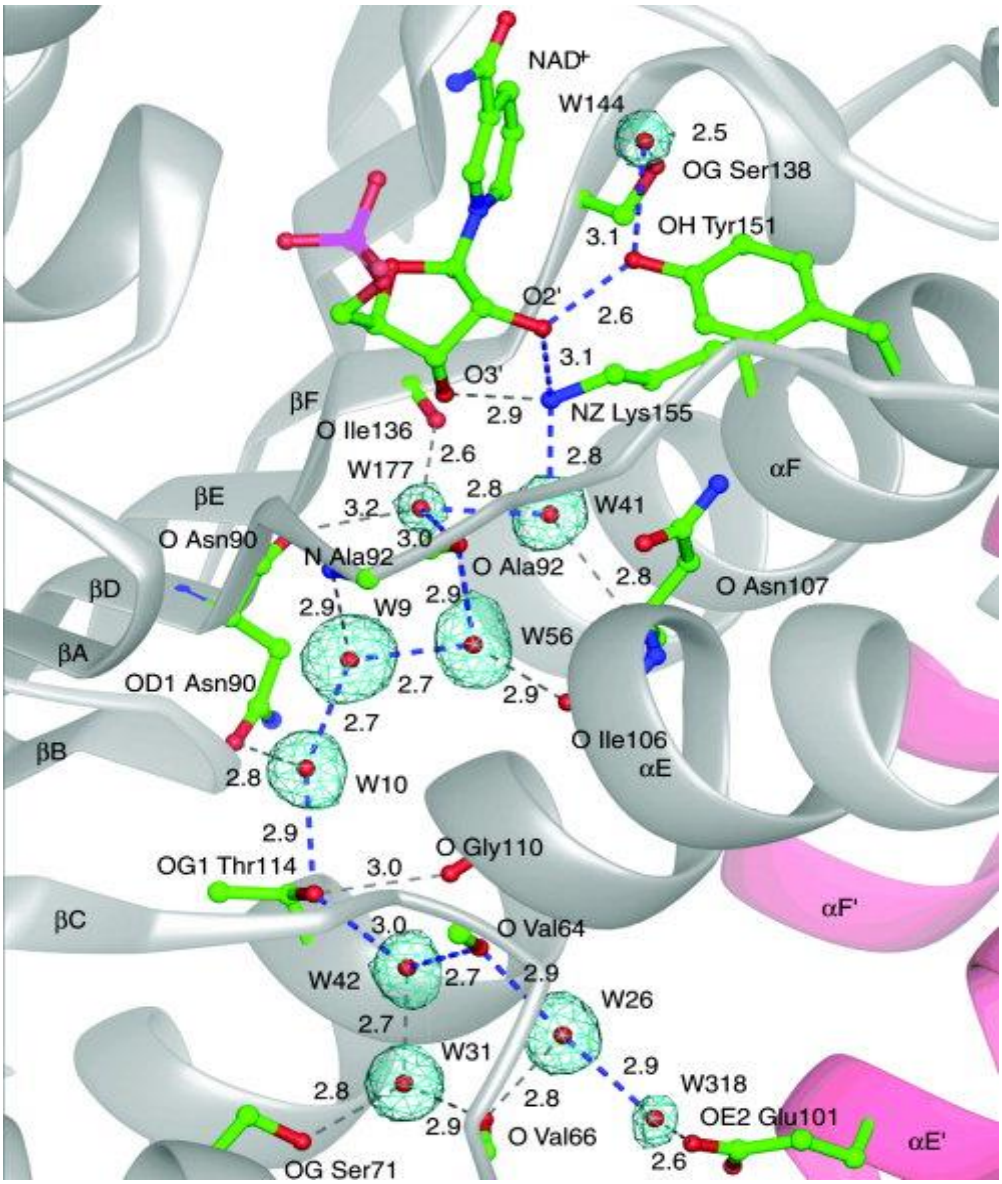


Figure 4.1.9 - the water chain for proton relay in alcohol dehydrogenase from *Drosophila lebanonensis*. NAD and important residues of the enzyme are shown in the stick representation, whilst the rest of the protein is shown in the ribbon representation. Water molecules are represented by red spheres. $2F_o$-F_c maps are shown for the water molecules. Image taken from the study of Koumanov et al.(289)

TDH exhibits flexibility at several distinct regions. Two disordered loop regions stand out most prominently. One includes residues Thr179-Ala185 and was designated 'Loop 1', and another is located closer to the N-terminus at residues Asp35-Asn60, and was designated 'Loop 2' (see Figure 3.1.15 and Figure 3.1.16). Other regions that appear to be flexible include Ser82-Thr100, Thr186-Pro205, Ser255-Ala273 and the last twenty or so residues leading up to the C-terminus. As previously discussed, the region showing the most significant variation in conformation is Loop 1. The role of this loop in L-threonine binding may be crucial in achieving the high specificity of the enzyme. Indeed, findings presented in this thesis show that TDH is very selective for catalysing the breakdown of L-threonine over other naturally-occurring amino acids. The other flexible regions and indicators of their disorder, such as the average B-factor, may also help to explain the dynamics of TDH during catalysis. Much of the conformational variability observed is in the catalytic domain of TDH. Furthermore, when the average residue B-factors of TDH in the 'open' conformation are compared with those in the 'closed' conformation, the disorder in the NAD-binding domain shows a relative decrease, whilst the disorder in the catalytic domain shows a relative increase, except for Loop 1, which is rigid, when closed (see Figure 3.1.19 and Figure 3.1.20). This suggests that disorder is maintained in the NAD-binding domain to allow the binding and release of the co-factor. As some correlation between the binding of ligands to the L-threonine binding site and closure of Loop 1 has been observed, it is possible that binding to the L-threonine binding site is associated with conformational changes in the catalytic domain of TDH. Hence, this may represent an induced fit binding mechanism.

X-ray crystallography and molecular dynamic studies of TDH from *Cupriavidus necator* have gathered more evidence that would fit this mechanism of ligand binding. Nakano et al. found that *Cn*TDH shows flexibility in the same areas as *Tb*TDH (residues 38-59, 77-87, 180-86 [analogous to loop 1] and the catalytic domain) in a crystal structure of *apo Cn*TDH. The same regions were found to be rigid in an L-threonine-bound structure. Molecular dynamics simulations indicated that the binding of NAD^+ induces a "rigidification" of flexible regions in the NAD-binding domain, whilst the flexible loop at residues 180-186 and the catalytic domain remain flexible to accommodate L-threonine binding(284). These findings are in some agreement with those presented in this thesis, which suggest that changes in the conformation of Loop 1 and the catalytic domain of TDH happen in response to L-threonine binding. The findings of Nakano et al. also support the existence of an induced fit mechanism of ligand binding for TDH. Loop 1, forms a "lid" over the L-threonine active site, something that would restrict L-threonine binding if it happened prior to its binding. Some researchers assert that all enzymes that form a lid over active sites must act by induced fit mechanisms(292); the findings on TDH structure would seem to illustrate this well. However, structures of TDH in the closed conformation have been obtained where only a solvent molecule, such as acetate, is bound. This suggests that, at least in the case of NAD-bound TDH, Loop 1 can take on both the open

and closed conformations, prior to binding L-threonine. The theory of conformational selection or conformational sampling takes a broader perspective on ligand binding by enzymes, whereby several conformations can be adopted by an enzyme prior to binding and catalysis. Only some of these conformations can enable ligand binding or catalysis(293,294). Thus, an induced fit mechanism may be seen as a subset of the wider process of conformational selection(295). In the case of TDH, several different combinations of conformations of the aforementioned flexible regions may be present among the population of TDH enzymes. The relative representation of each conformation may be influenced by the binding of ligands, but the fact that different conformations are sampled allows a greater variety of ligands to bind, beyond analogues of the natural substrates. This has clear implications for drug discovery and design which will be discussed further below.

4.1.3 Catalytic activity

Detailed kinetic studies of TDH, supported by the X-ray crystallographic data, indicate that the enzyme acts by a Bi Bi compulsory ordered mechanism of action (see Section 3.2.1). NAD^+ binds to the *apo* form of TDH prior to L-threonine binding. Then, following catalysis, AKB or aminoacetone dissociates from the enzyme before NADH is released. Evidence for such a mechanism was found for both *Tb*TDH and TDH from *Clostridium difficile* (*Cd*TDH). Double-reciprocal (Lineweaver-Burk) plots of the rate of catalysis observed at different combinations of substrate combinations showed a pattern that was indicative of a Bi Bi compulsory ordered mechanism, but not confirmative. This was particularly the case for the data regarding *Tb*TDH catalysis, where patterns of parallel lines were seen together with lines of decreasing slope gradients. Other potential mechanisms of action for TDH are Bi Bi random ordered and ping pong mechanisms. A ping pong mechanism is excluded by the fact there does not appear to be a group within the enzyme to which a hydride could be transferred to, before it is transferred to L-threonine. A Bi Bi random ordered reaction exhibits the same patterns as compulsory ordered reactions on Lineweaver-Burk plots, therefore, they can't be distinguished with these data alone. However, it was possible to gain X-ray structures of TDH that was bound to NAD, independently of L-threonine binding. Efforts to co-crystallise L-threonine and its diastereoisomer L-*allo*-threonine with TDH were only successful when NAD^+ or NADH were also present. Moreover, the uncompetitive inhibition shown towards NAD^+ by threonine binding site inhibitors (e.g. L-*allo*-threonine) confirms this hypothesis. Finally, researchers studying other TDHs have also made the conclusion that the enzyme acts by a compulsory ordered mechanism(197,291).

The K_M value and saturating concentration established for NAD^+ were approximately 1mM and 10mM, respectively. These values may appear high when compared to other NAD-dependent

enzymes. The K_M values for NAD^+ in other GalE-like TDHs have been reported to be as low as 0.1 mM for *Cn*TDH(126), 0.18 mM for mouse TDH(136), 0.18 mM for F*f*TDH(197) and 0.18mM for *Tv*TDH(125). However, several K_M values published for other GalE-like TDHs are higher and closer to those presented here. For chicken, goat and pig TDH, K_M values for NAD^+ have been reported as 0.98 mM(291), 1 mM(202), and 1 mM(201), respectively. Therefore, the NAD^+ value discovered for TDH from *T. brucei* is within the normal range to be expected for a GalE-like TDH. K_M values for NAD^+ of 0.2-0.25 mM were determined using methods similar to those employed here, but using TDH extracted from *T. brucei* and *T. brucei brucei* cell homogenates(139). Although substrate inhibition has not been described for other TDHs, competitive inhibition of NAD^+ by NADH has been reported frequently(197,282,291,296). Thus, the affinity of NADH for the NAD^+ binding site can explain the decrease in reaction velocity at NAD^+ concentrations above 10mM NAD^+: as the relative concentration of NADH compared to NAD^+ increases, product inhibition by NADH occurs. The affinity of TDH for NADH was also exploited to obtain a co-crystallised structure of NAD and L-threonine bound TDH, where turnover of L-threonine was prevented (see model TNADH4).

The values reported in the literature for the K_M and saturating concentrations of L-threonine are very similar amongst GalE-like TDHs. The saturating concentration of 30mM for *Tb*TDH found in this study is equal to the saturating concentration established in prior studies of *Tb*TDH(139,146). Similar results have also been found in studies of other GalE-like TDHs(125,126,202). However, the double-sigmoidal curve relating L-threonine concentration to reaction velocity (see Figure 3.2.8) has not been reported for any TDH to date. Bell & Turner studied TDH from several different organisms, including *Flavobacterium sp.* and *Bacillus sp,* and found that there was a sigmoidal relationship between reaction velocity and L-threonine concentration in all the enzymes tested(129). This indicates that TDH action is cooperative. However, cooperativity is something that has been little addressed in other studies of TDH. At present, the response to L-threonine concentration exhibited by TDH in these studies appears to be a unique feature.

There are a number of possible causes of double-sigmoidal kinetic behaviour. One possibility is that there are two different enzymes with different affinities for the substrate contributing to the observed catalytic activity(194,297). The SDS-PAGE-verified purification process that TDH underwent makes this possibility very unlikely. A more likely cause of the observed kinetic behaviour is an allosteric effect produced either by enzyme-substrate interactions or interactions between different TDH monomers (149,157–160). This possibility is strongly supported by the kinetic data. Size-exclusion chromatography results also strongly support the involvement of protein-protein interactions between TDH monomers, induced by substrate(s). Looking more closely at the kinetic data (see Figure 3.2.8), the two portions of the double-sigmoid show different apparent Hill coefficients, which implies that the degree of cooperativity is different

across the lower and upper ranges of L-threonine concentration. In the second portion of the curve, a greater degree of cooperativity is apparent than in the first part of the curve. The ligand-induced conversion of a receptor to an activated state has been shown to cause increases in the hill coefficient. Furthermore, such phenomena can manifest in complex, multi-phasic dose-response curves for receptors and enzymes(302). In light of this evidence, the catalytic behaviour of TDH alludes to the existence of two active TDH forms, with L-threonine inducing conversion between the two forms directly through binding in the active site or by acting at one or more allosteric sites. The differences between the two forms may be a difference in affinity or catalytic efficiency.

There are several different models to describe cooperative activity and they usually cite changes in binding affinity as a cause of such behaviour. Two of the most widely cited models of cooperativity are the Simple Sequential Interaction Model proposed by Koshland et al.(303) (KNF) and the Monod-Wyman-Changeux Concerted Transition or Symmetry model(304) (MWC). The KNF model describes a process by which binding of a ligand to a subunit of an enzyme oligomer causes a conformational change in that subunit, which in turn causes a conformational change in an adjacent monomer, increasing its affinity for the ligand. This continues sequentially throughout the molecule until all monomers are activated(194,303). The MWC model is somewhat simpler, in that it describes a concerted change of all enzyme monomers from a 'tense state', T, to a 'relaxed' state, R, upon ligand binding(194,304). When the kinetic data is considered in conjunction with the gel filtration studies, it raises the strong possibility that association and dissociation of TDH monomers is the root of the allosteric effect observed. Association/dissociation models have been used to describe kinetic behaviour of a number of enzymes in the past, and such models are also compatible with the KNF and MWC models of cooperativity. However, the MWC model has found more use in this respect, due to its classification of the entire enzyme complex as either a 'tense' state or a 'relaxed' state, which is also easier to relate to macroscopic rate constants obtained from assays such as those carried out in the studies presented in this thesis. The model of dissociating regulatory enzyme(297) is similar to the MWC model, but it describes different enzyme states in terms of their oligomeric forms. Some enzyme systems described by this model can show a reliance on the enzyme concentration. The size-exclusion chromatography data indicate that dissociation of the TDH dimer may occur. However, the influence of enzyme concentration would need to be established using further experiments, which might include using other techniques, as described later on in this discussion. The different models of cooperativity that could explain the cooperativity observed for TDH are summarised in Figure 4.1.10.

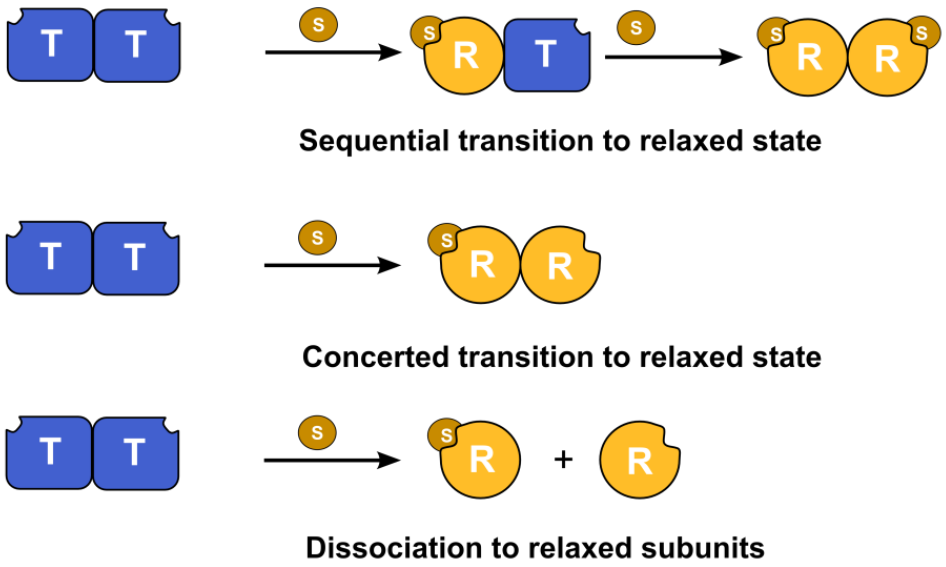


Figure 4.1.10 - models of cooperativity that may be applicable to TDH. From top to bottom, schematic representations of the KNF, MWC and dissociating regulatory enzyme models are shown. T = 'tense' state subunit; S = substrate; R = 'relaxed' state substrate.

Although the kinetic behaviour discovered in the presented studies of *Tb*TDH has not been reported for other TDH enzymes, similar behaviour has been reported for numerous other enzymes in the literature. These enzymes have been reported to dissociate, and the subunits have different activities or affinities for their substrates. In recent years, several of these proteins have been described by a concept known as the morpheein concept(305).

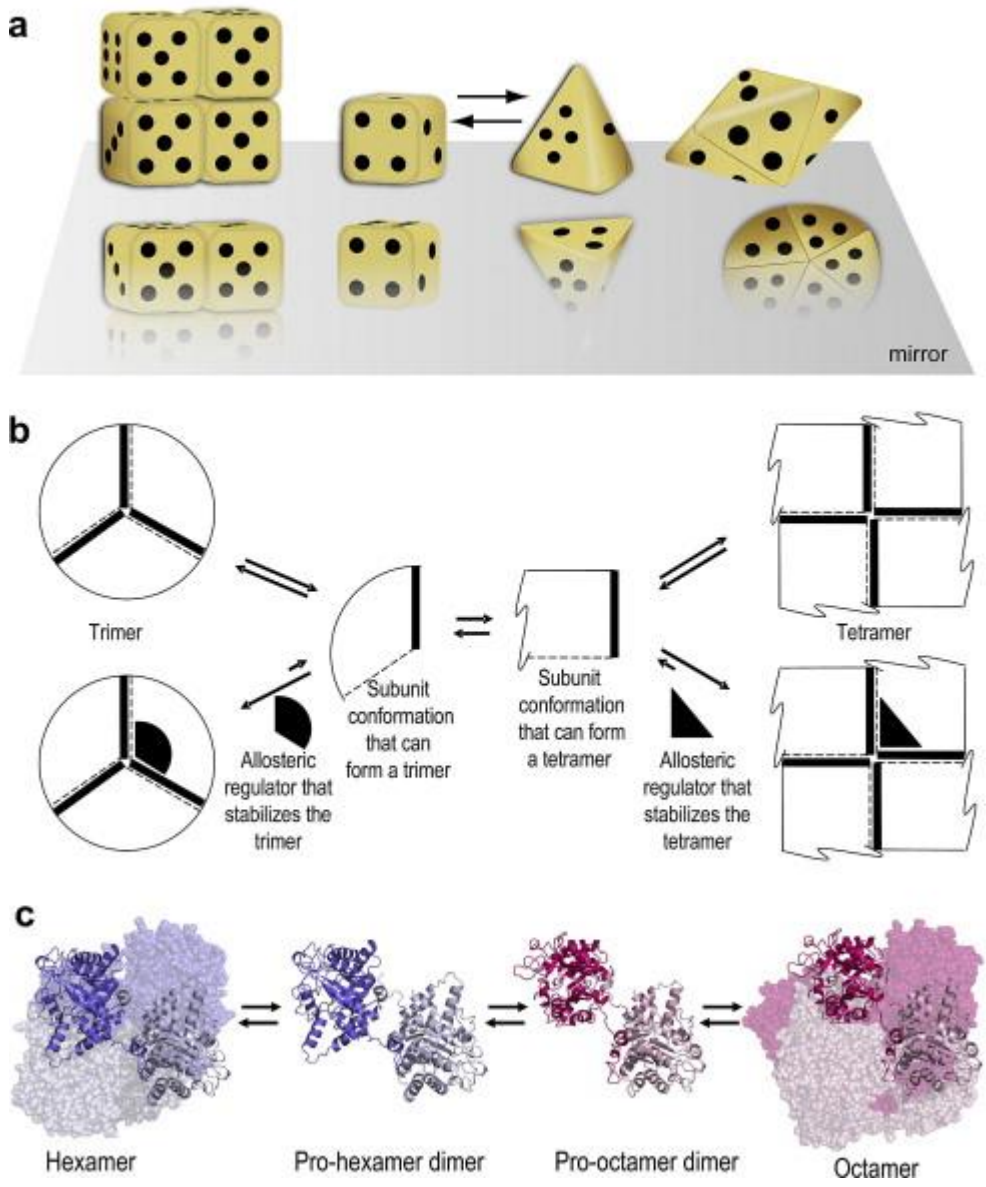


Figure 4.1.11 - three illustrations of the morpheein concept presented by Selwood and Jaffe(305). Image a uses the analogy of four- and three-sided dice to illustrate how different protein conformations form different oligomers. Image b schematically highlights the influence of allosteric regulators on the conformations of protein subunits and their oligomeric forms. Image c shows an example of crystallographically-determined morpheein behaviour by porphobilinogen synthase.

The morpheein concept describes enzymes that exist in multiple active oligomeric forms. The activities of the enzymes in different oligomeric forms are different, and this may lead to some of the peculiar kinetic relationships observed for these enzymes. In Figure 4.1.12, an example of the kinetic curve that can result from the presence of more than one morpheein form (or oligomeric form of an enzyme) is shown(305,306).

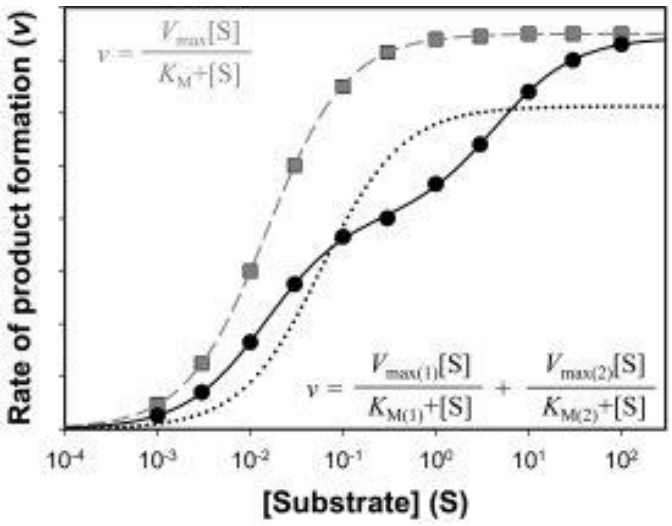


Figure 4.1.12 - example plots of substrate concentration as a function of product formation rate, exhibited by enzymes with classic Michaelis-Menten kinetics (black squares) and two co-existing, but kinetically distinct forms of an enzyme (black circles). The dotted line represents an attempt to fit the data represented by black circles to the Michaelis-Menten equation. Instead, the double-hyperbolic kinetics of that enzyme is modelled by an equation combining two Michaelis-Menten equations. This figure is reproduced from a publication by Lawrence and Jaffe(307).

A number of different enzymes that have been classified as morpheeins are dehydrogenases(305), suggesting that this may be a feature conserved across this class of enzymes. The enzymes described by the morpheein concept and by the dissociating regulatory enzyme concepts include several regulatory proteins. Given the role of TDH in several important cellular processes in *T. brucei*(138), and possibly several other organisms, it is a firm possibility that the apparent dissociation observed during size-exclusion chromatography could form an important part of the enzyme's physiological mechanism of action. For instance, it has been postulated that the increased TDH activity in procyclic *T. brucei* may occur in response to the greater availability of amino acids in the gut of the tsetse fly(138). Further support is given by the fact that the related enzyme GalE has been shown to dissociate from a dimer into two monomers. Brahma, Bhattacharyya and colleagues provide some of the most detailed studies of this effect in GalE from *Kluyveromyces fragilis*(308,309). Similar to TDH, GalE from *K. fragilis* is a homodimer of around 75kDa in mass that has one catalytic site per subunit. It has been demonstrated that, following partial digestion with trypsin, the monomeric subunits of GalE are catalytically active, independently of the other subunit(308). In a similar fashion to TDH, the native enzyme showed non-typical Michaelis-Menten kinetics, with a biphasic increase in catalytic velocity in response to increasing substrate (UDP-galactose) concentrations (see Figure 4.1.13). The kinetic behaviour of monomeric GalE (formed by treatment with *p*-chloromercuribenzoate) was shown to be distinct from the native dimeric form and was monophasic (see Figure 4.1.14). Finally, the selective inhibition of one active site confirmed that the biphasic kinetic behaviour was the result of two catalytically non-equivalent binding sites – a high-affinity site (active at lower substrate concentrations) and a low-affinity site (active at higher concentrations). The authors who published this data concluded that the activity of GalE from

K. fragilis was regulated by allostery between the substrate-binding sites of the two subunits(309). These findings are clearly in line with the findings of the kinetic behaviour of TDH which are presented herein: both enzymes show a biphasic response in catalytic activity towards the natural substrates. For TDH, there is evidence that substrate binding causes conformational changes, and these changes may include the dissociation of the dimer into two monomers. However, Brahma *et al.* demonstrated that the dimer was stable during catalysis, suggesting that fast association-dissociation between dimeric and monomeric forms does not occur in GalE. They also showed that a bound inhibitor, 5'-uridine monophosphate (5'-UMP), was responsible for some of the lag between the two phases of the kinetic curve, implicating an allosteric site that is distinct from the catalytic site(309).

These observations can be extrapolated to propose two other mechanisms for the regulation of TDH activity. Firstly, the binding status of each TDH subunit may influence the affinity of the second subunit. This situation is akin to a situation described by the KNF or MWC models. Secondly, an allosteric site that is distinct from the substrate binding sites may mediate the activity of TDH. This site may be bound by L-threonine and/or other ligands. As discussed previously, the MOIs exhibited by some inhibitors suggest that there may be binding sites on TDH where allosteric control can be exerted. For example, the *Mm*TDH inhibitor Qc1 appears to act uncompetitively in respect to NAD^+ (preferentially inhibits the TDH-NAD^+ complex), but acts noncompetitively in respect to L-threonine (the presence of L-threonine does not influence Qc1's activity). This suggests that there is a site other than the substrate binding sites that Qc1 interacts with to inhibit TDH activity. Given the available information on TDH, none of the proposed models can be confirmed, but the existence of any one regulatory mechanism – dissociation of the dimer, allostery between the two active sites or ligand binding to an allosteric site – need not be exclusive, and several mechanisms may coexist and contribute to the kinetic behaviour observed. Moreover, in this thesis multiple lines of evidence indicating the presence of these mechanisms are presented. Behaviour that has previously been demonstrated for other dissociating proteins, including some morpheeins, has been observed in size-exclusion chromatography and enzyme kinetic studies. The fact that the existence of two catalytically distinct oligomeric states has been demonstrated for the evolutionarily-related GalE also lends further support to the possibility that TDH may show similar behaviour. Analysis of X-ray crystallography structures have highlighted that the enzyme undergoes significant conformational changes in its tertiary structure (see Section 3.1.5), whilst the results of size-exclusion chromatography indicated changes in the quaternary structure of TDH (see Section 3.1.6). Together, these studies gave strong indications that such changes are ligand-induced, as is also indicated by the observance of cooperativity in enzyme catalysis. Future studies can be geared towards obtaining more detailed insights into these processes.

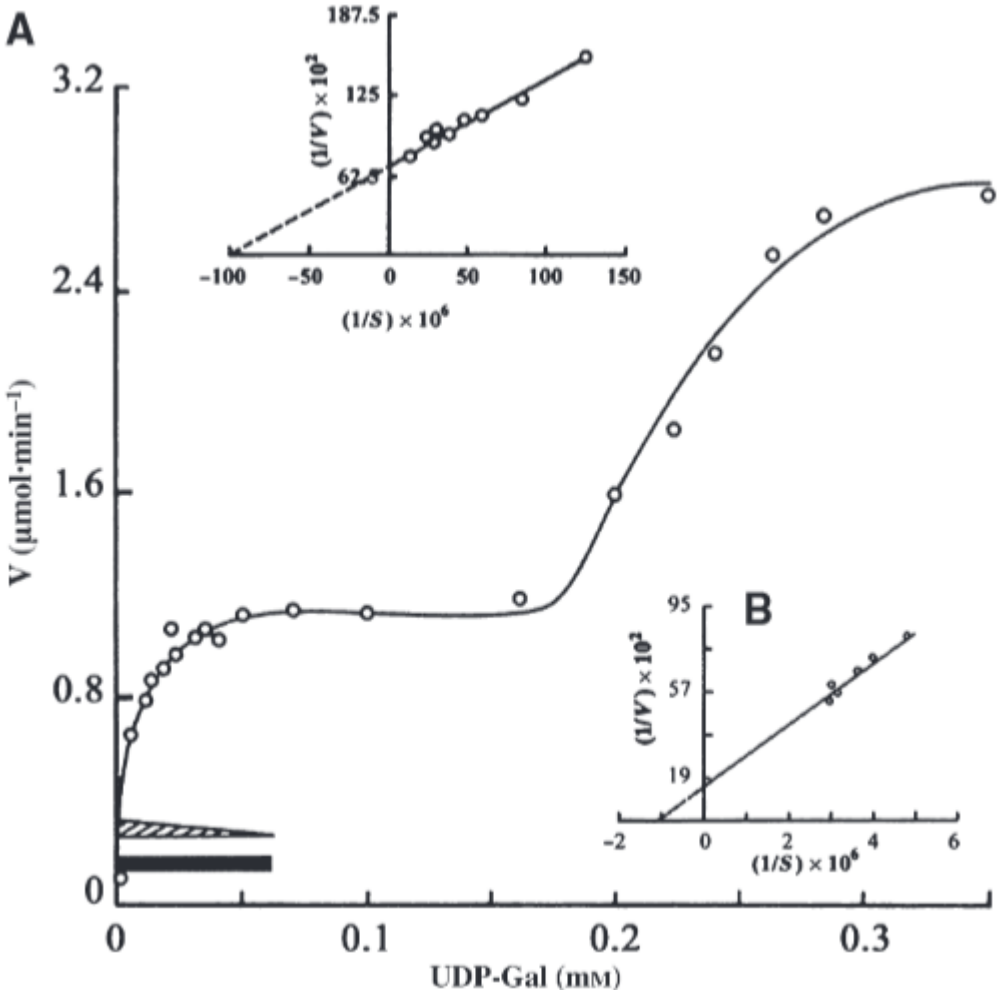


Figure 4.1.13 - kinetics of GalE from *K. fragilis* published by Brahma et al.(309). The largest plot shows the relationship between the velocity of enzyme catalysis and the concentration of UDP-galactose (UDP-Gal). The solid bar represents the presence of bound 5'UMP *ex vivo*, and the hatched bar above it represents the gradual disappearance of the initial lag in catalysis. A linear relationship could not be established between all data points on a single double-reciprocal plot. Therefore, the phase of activity at low substrate concentrations is plotted on one plot (A, top centre) and the points relating to the second phase of activity are included on a separate plot (B, bottom right).

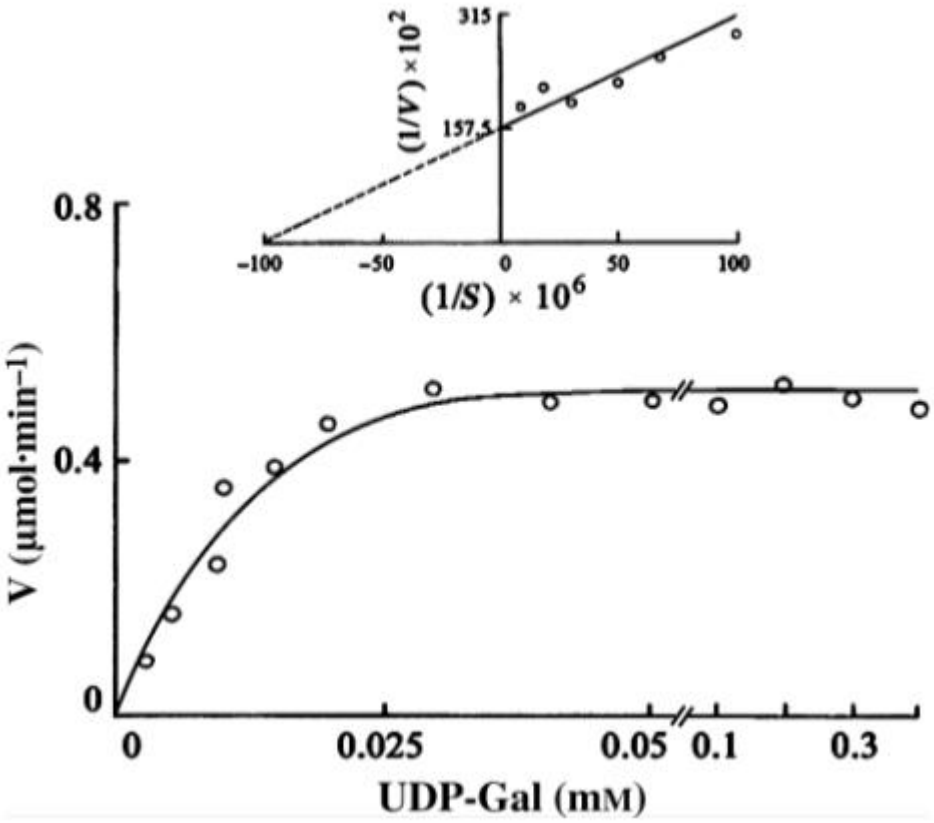


Figure 4.1.14 - a Michaelis-Menten plot of monomeric GalE from *K. fragilis* published by Brahma et al.(309). A double-reciprocal plot of the data is inset (top centre) The kinetics of the monomeric form are monophasic.

The modelling of TDH kinetic behaviour using equation 25, a modification of an equation proposed by Selwood, Jaffe and colleagues(273), was effective in describing the action of TDH as the model implies the co-existence of two different forms of the enzyme. A similar equation was used by Brahma and colleagues to model the biphasic kinetics of GalE from *K. fragilis*(309). As more information is gathered on the determinants of TDH kinetics, further parameters may be added to the equation, such as the contribution of each enzyme form to activity at different L-threonine concentrations and/or at specific TDH concentrations. This could also aid in the development of inhibitory equations based on this model. Unfortunately, at present the model does not reliably describe inhibition data better than normal Michaelis-Menten and Hill equations. The need to restrain several variables before fitting the equation means that some of the time-savings gained by using computer-based non-linear regression are lost.

There may be other ways to model TDH kinetics that have not been attempted in this work. The model that was used to describe the biphasic kinetics of TDH takes the assumption that each active site has a different affinity for L-threonine and/or a different catalytic efficiency. However, a limitation of this model and several others that are derived from Michaelis-Menten kinetic theories is that it does not delineate the differential effects that L-threonine may have on its own

affinity and on the catalytic efficiency of TDH. It is also implicit in the model described by equation 25 that the relationship between L-threonine concentration and rate is mediated by the same L-threonine binding interactions. It has been proposed that stimulatory and inhibitory effects should be described by equations that include separate modifier terms that separately describe a ligand's effects on substrate affinity and maximum reaction velocity(310). An example of such an equation is equation 37. The Michaelis-Menten equation,

$$v = \frac{[S]}{[S] + K_M} V_{max} \qquad \textbf{36}$$

which can calculate the catalytic velocity, v, at substrate concentration, [S], is expanded with terms for the effect of the modifier, X, on the substrate affinity, as indicated by the Michaelis constant, K_M, and the maximum velocity, V_{max}:

$$v = \frac{[S]}{[S]+K_{M1}-(K_{M1}-K_{M2})\left(\frac{[X]}{[X]+K_X}\right)} V_{max1} - (V_{max1} - V_{max2})\left(\frac{[X]}{[X]+K_X}\right) \qquad \textbf{37}$$

K_{M1} and K_{M2}, and V_{max1} and V_{max2} describe the K_M and the Vmax when a modifier is having no effect on them ('1')and when a modifier is exerting its maximum effect ('2'), respectively. The relationship between [X] and the degree of effect it has is described by the modifier constant, K_X. Although, such an equation would separately consider the effects of the modifier (L-threonine, in the case of TDH) on enzyme kinetics, more data would still be needed to confirm whether this model is appropriate – no information is available on whether the change in kinetics is due to changes in affinity, changes in catalytic efficiency or both. For this reason, this equation was not used to model TDH kinetics.

4.1.4 The relationship between TDH and KBL

The data from size-exclusion chromatography studies (see Section 3.1.6) support the classification of TDH as a dynamic dissociating protein and they indicate that it could potentially be a morpheein. However, no observations from X-ray crystallography and cross-linking experiments support the L-threonine-induced dissociation observed in size-exclusion studies. It is possible that L-threonine induces a large change in the shape of the TDH dimer, causing it to elute more slowly during size-exclusion chromatography. No gross changes in TDH conformation that could cause a drastic change in shape can be seen in X-ray crystallographic models, so this is unknown. There is actually more support for the dissociation of KBL and its morpheein-like behaviour, than there is for TDH dissociation. Cross-linking studies showed that the native quaternary structure of KBL was likely to be a dimer. In size-exclusion chromatography studies, three possible quaternary structures were suggested for KBL: a dimer, a monomer and an octamer. Furthermore, cross-linking by dimethyl suberimidate (DMS) of KBL monomers to form KBL dimers appeared to be inhibited in the presence of its substrate glycine (see Figure 3.1.35). It is unclear whether this effect was due to a preferential reaction between

DMS and glycine over KBL or if a higher oligomer such as an octamer was formed, but was not visible on the gel. The data on the dissociation of both TDH and KBL are inconclusive, as is the data that could support or reject the formation of a TDH-KBL complex. Both the pull-down assay and cross-linking studies do not support the existence of a stable complex between the two, but this does not exclude the formation of a transient complex. Efforts to co-crystallise TDH and KBL in a multi-enzyme complex also failed. However, in size-exclusion chromatography studies, it was clear that the elution of each enzyme was altered in the presence of the other. The enzymes co-eluted as monomers when they were injected onto the column in a single sample. There was also evidence that a small population of the enzymes co-eluted at volume that was suggestive of a higher molecular weight oligomer with a molecular mass of around 190kDa.

As previously stated, more data is required to conclusively demonstrate that TDH and KBL dissociate and form a complex. If TDH and KBL do form a multi-enzyme complex under certain conditions, the dissociating behaviour that has been suggested by data on both enzymes could be a part of the mechanism of complex formation. In this scenario the morpheein concept could be extended: homooligomeric proteins could dissociate, and change conformation to allow heterooligomeric complex formation.

4.1.5 Inhibition and stimulation of TDH

The inhibitory assays carried out with TDH and various other substances were effective in discovering new inhibitors and also in confirming the published reports of previously known inhibitors of TDH (see results in Section 3.2.3). L-*allo*-threonine, pyruvate and methylglyoxal are all reported to inhibit *Tb*TDH or TDH from other species. L-*allo*-threonine, a diastereoisomer of TDH, is perhaps the most widely investigated of these inhibitors. L-*allo*-threonine or DL-*allo*-threonine have been previously reported to inhibit *Tb*TDH(200), as well as TDH from *E. Coli*(119,311,312) and *S. aureus*(198). The K_i for L-*allo*-threonine determined here is 1.6mM, which is very similar to that determined by Klein and colleagues, 2.5mM(200). The mode of inhibition (MOI) determined by enzyme inhibition assays, which indicated that L-*allo*-threonine competes with L-threonine for its binding site, was confirmed by X-ray crystallography. The uncompetitive MOI shown towards NAD^+, is expected because of the high similarity of the inhibitor to L-threonine, which preferentially binds NAD-bound TDH. The inhibition of *Tb*TDH by L-*allo*-threonine, coupled with the lack of inhibition by and catalysis of D-*allo*-threonine by the enzyme, indicates the stereospecific nature of TDH. The L- orientation at the α-carbon is necessary for binding to TDH and this precludes the binding of the enantiomer of L-threonine, D-*allo*-threonine. In addition, the *R* configuration at the ß-carbon is necessary for catalysis. As a result, L-*allo*-threonine inhibits TDH(200). Interestingly, Tressel and colleagues reported that TDH from pig liver catabolises L-*allo*-threonine, albeit at a rate 10 times slower than the rate of L-threonine catabolism(201). The K_M for L-*allo*-threonine catalysis by TDH from pig liver was

calculated as 16mM, but no evidence exists to suggest that *Tb*TDH catalyses L-*allo*-threonine, either in these studies or in the literature.

Figure 4.1.15 - the structural formulas of L-threonine and analogous compounds, L-*allo*-threonine, pyruvate and methylglyoxal.

Pyruvate has been reported to inhibit TDH from *F. frigidimaris*(197) and *T. volcanium*(125). Like L-*allo*-threonine, it was shown to bind within the L-threonine binding site of TDH in an X-ray crystallographic model. Consequently, it shares the same MOI with L-*allo*-threonine, being competitive towards L-threonine, whilst acting uncompetitively towards NAD^+. Methylglyoxal, a close analogue of pyruvate, also inhibits TDH in a competitive manner, but is a noncompetitive inhibitor towards NAD^+. In contrast to the previous two inhibitors discussed, methylglyoxal inhibition was cooperative and apparent Hill coefficients of more than 3 were calculated from the experimental data. Methylglyoxal is reported to inhibit TDH from goat liver, and the curve relating inhibition to methylglyoxal concentration was sigmoidal, suggesting that the inhibition was cooperative and allosteric. Furthermore, treatment of goat liver TDH with trypsin differentially decreased the catalytic activity of the enzyme and the ability of methylglyoxal to inhibit it, indicating that methylglyoxal acts at an allosteric site and not at the L-threonine site(202). This would appear to contradict the finding that methylglyoxal inhibition is competitive towards L-threonine. Indeed, methylglyoxal was not observed in the L-threonine binding site in X-ray crystallographic models. Methylglyoxal is produced as a toxic by-product of several physiologic pathways, most notably, glycolysis and is known to react with arginine and lysine residues on proteins (see Figure 4.1.16). Therefore, it is plausible that the allosteric inhibition seen is due to the modification of key arginine and lysine residues, such as those at the dimerisation interface. This would also explain the observations of Ray and Ray(202), that digestion of TDH from goat liver with trypsin, which selectively cleaves at the C-end of arginine or lysine residues(313), reduces the inhibition of methylglyoxal more than it does the catalytic rate of TDH. However, there are still questions as to why methylglyoxal shows a competitive MOI towards L-threonine. The only arginine or lysine residue that interacts directly with the substrates is Lys148, but binding to this residue would have manifested in an MOI that was competitive towards NAD^+. Sigmoidal inhibition curves can also be a sign of covalent modification, which would be explained by methylglyoxal's known action. It is possible that methylglyoxal binds reversibly to the L-threonine binding site, but also inhibits TDH by covalently modifying important residues. Unfortunately, X-ray crystallography has not revealed any of the sites of methylglyoxal's action.

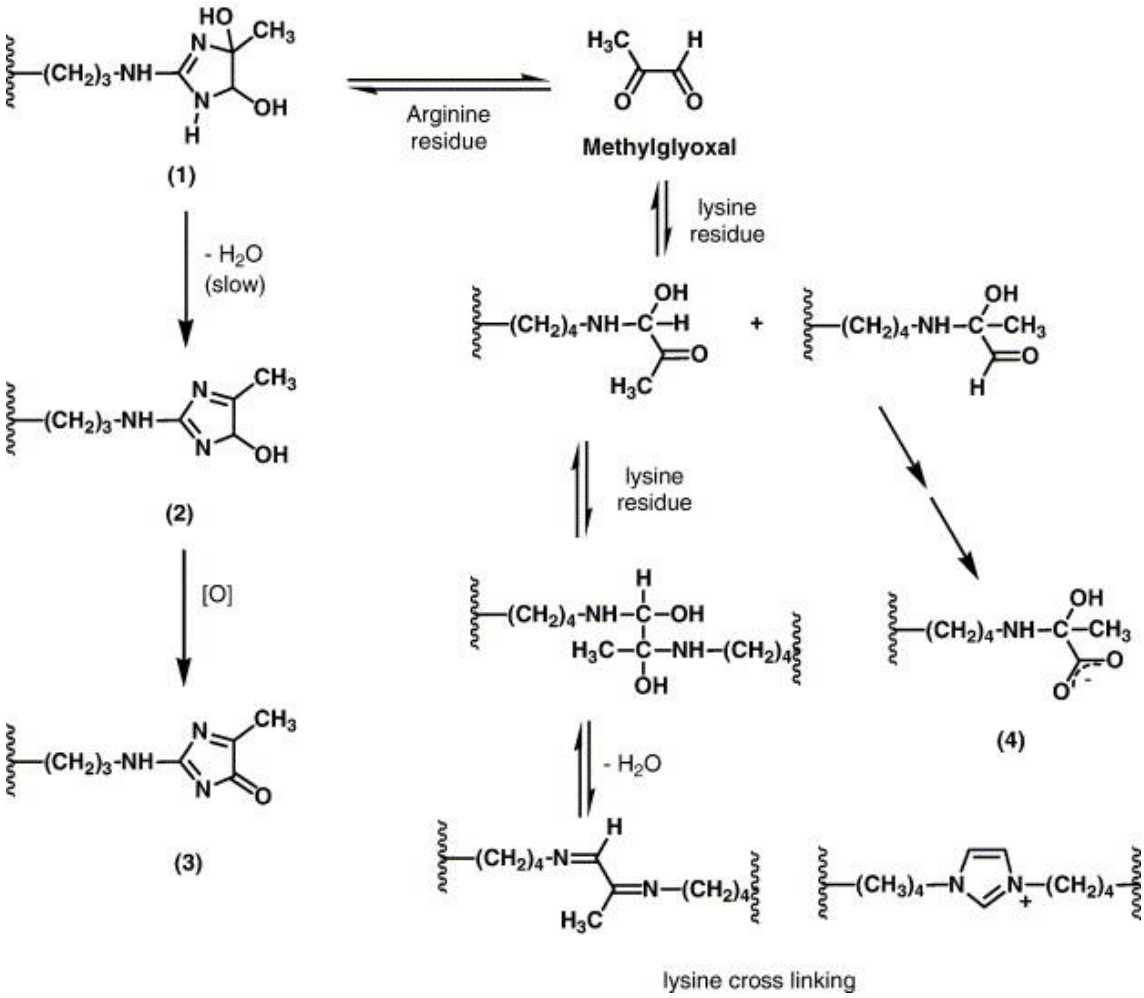


Figure 4.1.16 - the reaction of methylglyoxal with lysine and arginine side chains. Methylglyoxal is a reactive aldehyde that forms covalent adducts with the affected residues. Reaction of lysine adducts with a second lysine residue can lead to cross-linking. This figure was published by Bechara *et al*.(314).

There is a possibility that the inhibition of *Tb*TDH by pyruvate and methylglyoxal has some physiological significance in *Trypanosomes*. Pyruvate is a by-product of glycolysis, and it is also the substrate for pyruvate dehydrogenase, which is essential in the alternative pathway for the essential production of acetyl-CoA(143,144,149). Methylglyoxal is also a by-product of glycolysis in trypanosomes(149). Therefore, inhibition of TDH by pyruvate and methylglyoxal may actually be a feedback mechanism between the two pathways, whereby the TDH pathway is inhibited when the availability of glucose is high. In the goat, methylglyoxal is also produced by an amine oxidase from aminoacetone, the AKB breakdown product. Therefore methylglyoxal may act as a form of negative feedback to L-threonine catabolism by TDH(202).

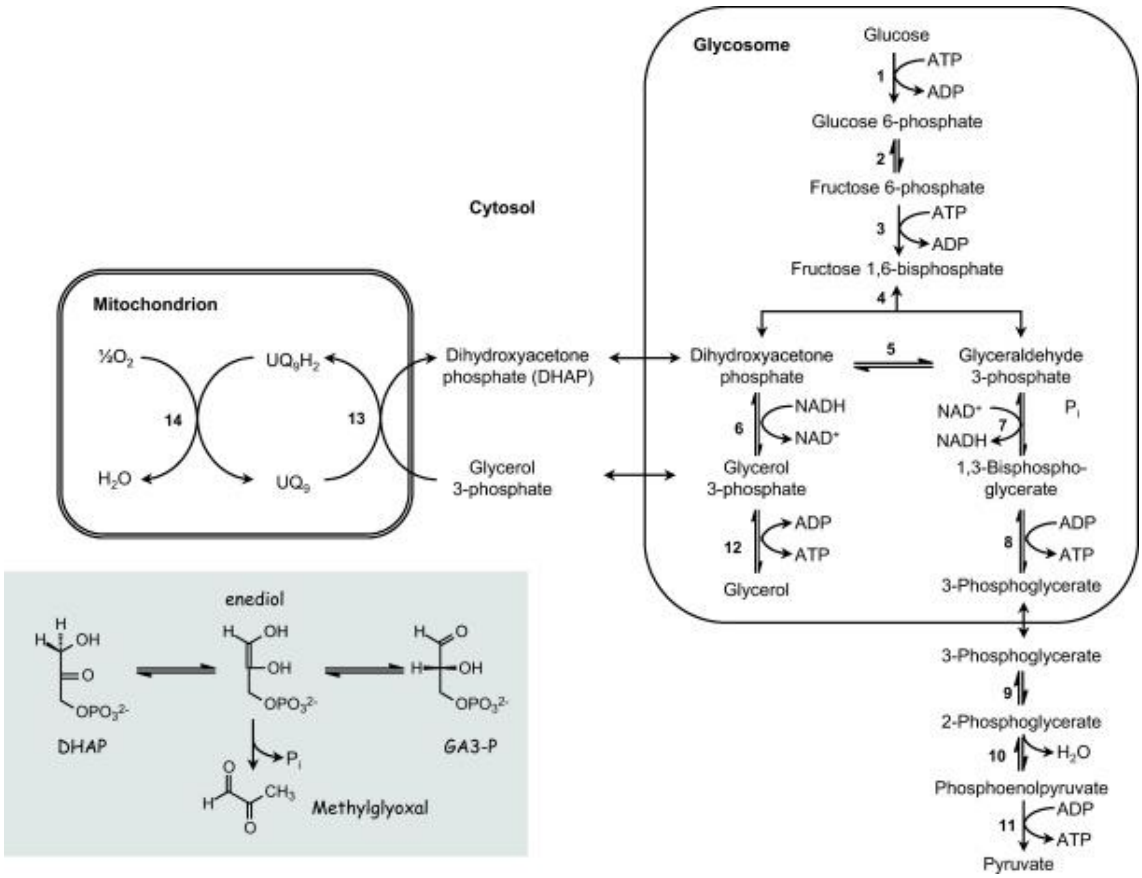


Figure 4.1.17 - the production of methylglyoxal as a by-product of the glycolytic pathway in trypanosomes. The shaded box illustrates production of methylglyoxal from dihydroxyacetone phosphate (DHAP). This figure was published by Wyllie and Fairlamb(149).

Tetraethyl thiuram disulphide (TETD) is one of the most potent known inhibitors of TDH. It was the trypanocidal activity of this compound that identified TDH as a potential drug target for human African trypanosomiasis (HAT)(138,139). In this study, TETD was found to be a time-dependent inhibitor of TDH. An analysis of the relationship between the rate of inactivation of TDH and the concentration of TETD indicated that the mechanism of inhibition was irreversible. This is in line with the findings of Cross, Linstead and colleagues that the inhibition of TDH by TETD was progressive over time, and that this inhibition was caused by covalent action. They also found that other sulfhydryl-reactive agents, including *p*-choromercuribenzoate, iodoacetamide and N-ethylmaleimide, inhibited TDH, indicating that there is an essential sulfhydryl residue in TDH(139). These findings were corroborated in this study: iodoacetamide and N-ethylmaleimide showed a time-dependent inhibition of TDH in a similar manner to TETD. A likely candidate for this essential sulfhydryl residue is Cys56, which was shown in TDH structure models to be involved in binding NAD^+. Unfortunately, disulfiram and N-ethylmaleimide were not successfully co-crystallised with TDH, so it has not been possible to confirm this as the site of action of TETD and other sulfhydryl-reactive agents.

Overall, the enzyme inhibitory assays have been a valuable tool for characterising the different MOIs that compounds may show against TDH. This knowledge is valuable in enzyme inhibitor

design because certain binding modes may be preferred. For example, uncompetitive inhibitors can be useful as drugs because their inhibition causes the accumulation of substrate, thereby encouraging the formation of the enzyme-substrate complex for which they have greater affinity. Time-dependent inhibitors can also be attractive drug candidates as their longer residence time when bound to the target can reduce their likelihood of interacting with non-target proteins, potentially reducing the incidence of side-effects(315,316).

TDH appears to be very selective for L-threonine as a substrate, although it may have some affinity for other standard amino acids, which were found to cause marginal inhibition at high concentrations (see Table 3.2.1). L-cysteine (60 mM) inhibited TDH by approximately 40%. Hopefully, X-ray crystallography will be able reveal the mechanism of action of this inhibitor. It is plausible that it binds within the L-threonine binding site, or it may react with an important sulfhydryl residue in the enzyme. Another unexpected observation was that TDH activity was increased in the presence of L-glutamine. This amino acid does not appear to be a substrate for TDH, so it may increase activity by binding to an allosteric site, or via another mechanism. Once again, X-ray crystallographic studies will be pursued to determine the cause of this effect.

Despite the fact that no specific monovalent cation binding sites were observed in X-ray crystal structures, certain monovalent cations were found to stimulate the activity of TDH to more than two-fold the original activity (see Section 3.2.2). In the presence of ammonium (NH_4^+), potassium (K^+) and rubidium (Rb^+) ions, the activity and thermostability of TDH appeared to be increased. Stimulation of TDH by K^+ has been reported for TDH from *S. aureus*(198,317,318), as well as for TDH from *T. brucei*(139). Stabilisation of enzymes and stimulation of enzyme activity by ions can occur through specific effects (e.g. involvement in substrate binding) or through physical effects. Ions can have a physical effect on enzymes through stabilisation or destabilisation of the structure of water. Kosmotropes stabilise water, whilst chaotropes destabilise water. Kosmotropic anions and Chaotropic cations, such as K^+, NH_4^+ and Rb^+ stabilise enzymes(319–321). If there is indeed no specific binding site for monovalent cations on TDH, then the effects observed may be physically mediated effects.

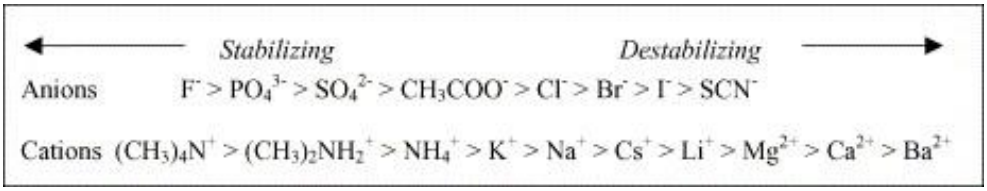


Figure 4.1.18 - the Hofmeister series, showing the order of enzyme stabilising and destabilising effects caused by different ions. This series was published by Zhao.(319)

The order of stimulating activity, NH_4^+,K^+<Rb^+ is similar to the order of ion chaotropism derived from the Hofmeister series(320–322). The effects of different ions on water structure are related to their atomic radii and charge(319). Thus, identifying the true cause of an ion's effect on an enzyme is made more difficult by the fact that the selectivity of an enzyme for a particular ion is also related to the ion's atomic radius and charge(323). More structural and biochemical studies will be required to understand the nature of the relationship between the activating monovalent cations and TDH. Crystallisation of TDH in solutions containing high concentrations of K^+ or Rb^+ failed, but a successful strategy that should be tried is to crystallise the protein of interest with Thallium ions(324,325). These ions have a high electron density, so will be easily distinguished from water molecules and other solvent molecules in electron density maps.

When the activity of TDH was measured in the presence of divalent cations, a very different effect to that seen with monovalent cations was observed. Zinc (Zn^{2+}), cadmium (Cd^{2+}), copper (Cu^{2+}), aluminium (Al^{2+}), cobalt (Co^{2+}), calcium (Ca^{2+}) and iron ions (Fe^{2+} and Fe^{3+}) all inhibited TDH activity. This contrasts with the forms of TDH from *E. coli,* and *Pyrococcus* spp., which require divalent cations, such as Zn^{2+}, Mn^{2+} and Cd^{2+} and Co^{2+}(121,123,326). Manganese ions had no effect on *Tb*TDH and the other ions important in other forms of TDH inhibited it. The inhibition of TDH by heavy metal ions, such as silver (Ag^+) and mercury (Hg^{2+}) provides further evidence that there is an essential sulfhydryl residue in the enzyme, and this is consistent with findings from other GalE-like TDHs, from *F. frigidimaris* and *C. necator*(126,197). Inhibition of *Tb*TDH by Ag^+, Hg^{2+} and Cu^{2+} has also been previously reported(139).

4.1.6 Evaluation of Investigative Methods

The use of recombinant DNA for the production of TDH and KBL enabled the development of a repeatable and efficient method of protein production. The yields were sufficient that enough protein could be obtained to carry out the necessary experiments over a reasonable period of time. As the initial focus of the studies conducted here was X-ray crystallography, the main aim of the protein purification methods was to obtain a pure enough sample to allow successful crystallisation. SDS-PAGE was used to measure protein purity at each stage. Even in concentrated samples, stained TDH and KBL bands were visible, whilst no other bands were visible after the affinity chromatography step. The purity of such samples was confirmed by size-exclusion chromatography experiments, which only detected protein elution under a single peak. Although often carried out during the purification of enzymes, specific activity (often reported as units of substrate catalysed per mg of protein, per minute) was not used as a measure of protein purity. As there is no reference value of specific activity for pure TDH or KBL from *T. brucei*, specific activity could not be used to estimate absolute purity. However, for future studies of TDH, it would be useful to use specific activity as a reliable way to determine

the relative purity of the sample at each stage of the protein purification process. The reliability of SDS-PAGE as a way to measure could be improved by applying a densitometric analysis to SDS-PAGE gels.

In the pursuit of structural and functional information on TDH and KBL, the use of multiple orthogonal assays to gather supportive data brought many advantages. For instance, to understand the structure of the enzymes, X-ray crystallography provided information on primary, secondary, tertiary and quaternary structure. This information was enriched by data from SDS-PAGE, which confirmed the primary structure, and size-exclusion chromatography and cross-linking studies confirmed the quaternary structure. In the same manner, the functional information gained from enzyme assays with TDH was enriched by the ligand-bound structures obtained by X-ray crystallography. A primary motivation of this work was to initiate drug design, so the complementary information on inhibition that was gained from enzyme assays and X-ray crystallography is highly valuable. Overall, the data from the various methods involved allowed the identification of relationships between structure and function. Most notably, enzyme assays, X-ray crystallography, size-exclusion chromatography and cross-linking all produced data that indicated a relationship between protein-ligand interactions and protein structure. Furthermore, this information was able to inform the design of virtual screening and *in vitro* screening assays (see Sections 4.2.1 and 4.2.2 for a more detailed discussion). The consideration that different inhibitory compounds may bind to different conformational states of TDH, which may be present under different conditions, may have improved the success of screening and certainly allowed for the identification of inhibitors that were both chemically diverse and mechanistically diverse. In spite of all of the advantages brought by the approach used in this thesis, the information on TDH structure, function and inhibition remains incomplete. One reason for this is that the methods employed all have limitations, particularly in regards to what they are able to detect. For example, there may be protein-protein interactions that are either difficult to detect or difficult to gain detailed information on using X-ray crystallography and size-exclusion chromatography. Such a lack of information also limits the extent to which data can be analysed. For example, having limited insight into the relationship between TDH-ligand interactions and TDH function made it more difficult to model TDH function. The methods are evaluated in more detail below.

X-ray crystallography is often considered the definitive method for the determination of macromolecular structure. In this respect, it has found wide application in the elucidation of physiological processes and in the design of new drugs(15,17). The presented investigation provides yet another example of this application. Not only has the 3D configuration of TDH been determined, but the way in which the configuration of the enzyme varies has also been highlighted. The absolute structure has been used to characterise the enzyme and to confirm

its place as a member of the SDR superfamily and as a GalE-like TDH. Key conformational changes that may be related to ligand binding and catalysis also give insights into the enzyme's mechanism of action. Furthermore, the crystallography studies have identified the binding relationships between TDH and its natural substrates, NAD^+ and L-threonine. This is not only valuable from the perspective of investigating the enzyme's mechanism of action, but it will also be valuable in the design of new inhibitors of TDH. Indeed, X-ray structures of TDH bound to inhibitors have allowed an identification of the determinants of ligand binding, catalysis and inhibition.

To obtain information on how TDH binds its ligands, two strategies were used: the enzyme was co-crystallised with ligands, or crystals of the enzyme were soaked in a solution containing ligands. Only co-crystallisation proved to be successful. This method may also be preferred because, as ligand binding may be associated with significant conformational changes of a protein, binding of a ligand to a crystallised protein may cause large changes in the unit cells of proteins and may cause the crystal to re-dissolve. For ligands that were unsuccessfully co-crystallised, crystal soaking may be required. During crystal soaking experiments, TDH crystals were soaked in ligand solutions for up to one hour. However, more time may be required, particularly for low-solubility or slow-binding ligands, for this to be successful. The solubility of a ligand can limit the success of co-crystallisation or crystal soaking in obtaining enzyme-ligand complexes. Several variables need to be optimised, including the enzyme concentration, ligand concentration, soaking time and solvents. Another method that can improve the success of these approaches is the use of additives to improve ligand solubility(327). Further optimisation of co-crystallisation and crystal soaking experiments should be employed in the future to obtain ligand-bound TDH structures.

The use of other biochemical techniques for structural and functional characterisation of TDH was complimentary to X-ray crystallography, but may also have identified features of the enzyme that were not captured in crystallographic structure models. Size-exclusion chromatography is an example of this because it was able to provide evidence that TDH and KBL may exist in multiple oligomeric forms, something that was not captured by crystallographic models. A limitation of the size-exclusion studies carried out here is that a small change in elution volume amounted to a significant change in the predicted molecular weight of the sample. For example, a change in elution volume of 2-3ml could change the prediction of the quaternary structure of TDH or KBL from a dimer to a monomer. Nevertheless, there were samples that eluted at volumes that would make quaternary structure predictions ambiguous. The size-exclusion data sometimes did appear to be able to distinguish between different oligomeric forms, but more confidence would be gained by confirming the findings using other techniques. As discussed above, cross-linking studies served a similar purpose to size-exclusion studies. They corroborated the evidence for the dimerisation of TDH and KBL, something which has found agreement across three methods of structure determination used in

this study. They also provided further evidence of the dissociation of KBL, although this needs to be confirmed in an optimised cross-linking assay.

The findings relating to the formation of a multi-enzyme complex between TDH and KBL are conflicting. Size-exclusion studies clearly showed that that the quaternary structures of both enzymes were altered in each other's presence, but X-ray crystallography studies, cross-linking experiments and the pull-down assay all failed to find evidence to support this. It is possible that, if a transient complex is formed, none of the techniques used is able to measure this. Alternatively, it may be that the assays need to be optimised. It has been observed that ligands affect the conformation and quaternary structure of TDH and KBL, thus the presence and concentrations of different ligands can be varied. The pull-down assay was carried out in the absence of any ligands, so the effects of different ligands, particularly NAD^+, L-threonine and glycine, on the assay result should be tested.

There are other biological techniques that may be used to answer the outstanding questions on TDH structure and function. Dynamic light scattering (DLiS) measures the fluctuations in the absorbance of scattered light over time to calculate the hydrodynamic radii of particles in solution or suspension. From this, the molecular weights of proteins can be calculated. In contrast to chromatographic methods, DLiS is able to predict the molecular weight of molecules without potentially disturbing any protein-protein interactions, as in affinity chromatography. DLiS has been successfully used to identify and characterise protein-protein interactions in both permanent and transient protein complexes(328–330). The different variables that can affect protein-protein interactions (e.g. pH, temperature, ionic strength) can easily be varied in DLiS experiments, without affecting the functionality of the assay. Furthermore, DLiS has been successfully used to monitor the effect of small molecule ligands on oligomerisation and complex formation, so this could be a very useful technique for studying TDH and KBL(329). Isothermal titration calorimetry (ITC), a powerful tool for determining the thermodynamics of ligand-macromolecule interactions can also be used for the determination of protein-protein interactions. Specifically, the binding enthalpy, stoichiometry of binding and the equilibrium binding constant can be determined for a particular process(331). Like DLiS, ITC has been used to characterise transient protein-protein interactions(332).

Analytical ultracentrifugation (AUC) can produce very useful data by studying the different migration of proteins in solution, on application of a centrifugal field. This migration is influenced by the size and shape of the molecules being studied, so by understanding the relationship between gravitational force, particle mass and friction, the molecular weights of protein oligomers can be identified. There are two methods of AUC that can be used to study oligomers and multi-protein complexes: sedimentation velocity, which involves studying the different rates of migration of particles to determine molecular weight, and sedimentation equilibrium, which involves studying the concentration distribution of protein species when the forces of sedimentation and diffusion are in equilibrium(333,334). As AUC can be used to model molecular mass, hydrodynamic shape and stoichiometry of protein complexes, it could

potentially be very useful for studying the self-association of TDH and KBL monomers, as well as the formation of a TDH-KBL complex. Modern advances in AUC techniques have also allowed the detection and characterisation of dynamic processes, such as transient protein interactions. This would not only help to investigate any interactions between TDH and KBL, but it would also enable the study of the response of these proteins to ligand-binding.

Nuclear magnetic resonance spectroscopy (NMR) is another technique that could be used to answer several of the outstanding research questions in this work. NMR measures the energy released following changes in the spin states of nuclei induced by an applied magnetic field. The signals recorded are characteristic of the local environment of the nuclei and the method is able to detect effects caused by the coupling of spin states between geometrically close nuclei, which allows one to deduce the structure(s) of the substance of interest. NMR is the second most common method for biomolecular structure determination, after X-ray crystallography, and has the advantage over crystallography that experiments can be carried out in solution, rather than in the crystalline state. Thus, NMR studies can be complementary to X-ray crystallographic studies, or they can be used to gain insight where it is not possible to with X-ray crystallography(335). For the further study of TDH, NMR would be useful for capturing the structure of TDH in the oligomeric states that have been suggested by some of the data, but that have not been captured by X-ray crystallography. In particular, if TDH does indeed form a complex with KBL, which has not been successfully crystallised thus far, NMR could be used to confirm the existence of the complex and determine its structure. The protein cyanovirin-N is an example of a protein that has been shown to form dimers by X-ray crystallography and has been shown to exist in monomeric form by NMR. Barrientos *et al.* were able to confirm using NMR that the protein exists in two oligomeric states in solution, depending on experimental conditions(336). It has also been demonstrated that NMR can capture weak, short-lived protein-protein interactions(337), which would be useful for studying a transient multi-enzyme complex (assuming a TDH-KBL complex has these characteristics). Like X-ray crystallography, NMR can be used to probe ligand interactions. As alluded to above, the ability to observe some ligand interactions by X-ray crystallography may be limited by the fact that the conformation of the protein that binds a particular ligand may not be crystallised as readily as other conformations. This may have been the reason why it was difficult to co-crystallise TDH with some ligands, or to observe TDH-ligand interactions after using the crystal-soaking method. The kinetic behaviour of TDH is also suggestive of allosteric regulation by L-threonine and other ligands. However, no allosteric sites have been identified. NMR has been shown to be effective, not only for identifying multiple binding sites on proteins, but in combination with ITC, it has been used to obtain binding-site specific binding affinities, enabling a detailed description of the cooperative binding of ligands to human ileal bile acid binding protein(338). Thus, NMR may be a highly effective way to study TDH and its interactions with small molecules more deeply.

The differential absorption of NAD^+ and NADH is advantageous for the study of many NAD-dependent enzymes. The absorbance assays used in the experiments presented here were reasonably simple to design and scale up. A drawback of measuring absorbance to monitor enzyme catalysis is that several other compounds absorb in the region of TDH. This drawback was avoided by use of a kinetic assay, where absorbance measurements were made at several time-points, rather than an end-point assay, which would be more difficult to design to overcome this limitation. NADH is also a fluorescent molecule and problems of assay interference by absorbing compounds can be avoided by the use of fluorescence as a signal for monitoring enzyme catalysis. Fluorescence sensors are also more sensitive, so reagents can be spared as lower quantities are required(194). Fluorescence capabilities were not available during the entire course of the studies of TDH, and switching to fluorescence measurements may have required a significant reformatting of the assay. Therefore, the decision was made to use absorbance measurements and design the enzyme assays in an appropriate way to achieve the desired outcomes. The availability of the Omega plate readers for inhibitory assays and *in vitro* screening allowed the collection of more data, and subsequently a more thorough analysis before drawing conclusions. Lineweaver-Burk plots are a useful way of visually characterising enzyme mechanism of action and inhibitor MOIs, but having more data allowed the determination of these features using non-linear regression, which is now considered to be the preferred method(194,203).

Whilst biochemical techniques such as ITC can be useful and accurate in determining the properties of protein-ligand binding interactions, the use of enzyme assays as the principal tool for investigating TDH-ligand interactions is more appropriate as they are functional assays. What is of greatest interest is the effect of a ligand on the enzyme's function, and other biochemical techniques, which may give more accurate information on binding parameters, are less effective in providing information on the inhibition of function. The design of MOI assays, that tested the effects of varying the concentration of one substrate, at a fixed concentration of the second substrate, effectively yielded the data required for MOI determination. It can be argued that inhibiting compounds could have been spared by fixing the substrate concentration at a concentration near to its K_M, rather than at the saturating concentrations. Moreover, when the inhibitor was a competitive inhibitor of the substrate at a fixed concentration, the saturating concentration of substrate may have meant that the inhibition observed was low, making it more difficult to determine the inhibitor's effects in relation to the second substrate. The decision was made to carry out the assays at saturating substrate concentrations because it increased the size of the signal (absorbance, due to a faster rate of NAD turnover), allowing inhibitory effects to be accurately measured across a broad range of concentrations of the second substrate.

The binding relationships of the natural substrates of TDH and the MOIs of known TDH inhibitors were confirmed with X-ray crystallography, highlighting the effectiveness of the

enzyme assays. However, MOI determination and parameter estimation (e.g. K_i) was more difficult when studying the stronger inhibitors of TDH. The accuracy of the MOI predictions for some inhibitors is also yet to be confirmed by X-ray crystallography.

One potential source of error during inhibitory assay analysis was that the models for inhibition were sub-optimal for modelling the interaction between inhibitors and L-threonine, due to the atypical kinetics of TDH in relation to L-threonine. To improve on this situation, more knowledge is needed on the underlying processes causing the biphasic kinetics observed. The effects that inhibitors may have on this kinetic signature will also need to be discovered before more accurate models can be developed. Nevertheless, the current models used were able to differentiate between different MOIs, some of which were confirmed with crystallographic models. For the purposes of calculating the K_i of inhibitors, different forms of the Cheng-Prusoff equation can be used to determine the K_i from the IC_{50}. The correct form of the equation can be selected on the basis of the observed MOI.

A further complication of determining the MOI arises from the fact that the enzyme shows signs of cooperative behaviour and potential allosteric regulation. Some of the inhibitors studied showed signs of cooperative inhibition, which could be a result of a number of different processes: binding to an allosteric site; disruption of the cooperative binding of L-threonine; inhibitor-induced conformational changes. To model the inhibition of the compounds showing cooperative inhibition, standard inhibitory equations were modified by incorporating the Hill coefficient (see equations 29-34). Such equations have been used by other researchers to model cooperative inhibition(275,339). In modifying the equations in this way, there is a risk that the model is manipulated to allow it to fit to data. To ensure that this was not the case, the Aikake information criterion (AIC) and the F-test were employed.

One additional difficulty in the study of TDH enzyme inhibitors was that several of the inhibitors were time-dependent. Therefore, the screening assays and all the assays for MOI determination were lengthy and required pre-incubation of the inhibitors with TDH. Two assays were used for the investigation of time-dependent inhibitors. For studies of TETD, the activity of TDH was measured after it had been incubated with the inhibitor for different periods of time. This was useful, as the rate of inactivation could be measured which was then used to determine the mechanism of inhibition of TETD. However, this assay was quite cumbersome and required that the timing of initiation of each assay was consistent between repeat experiments. In the assays to determine the MOI of TETD in relation to NAD^+ and L-threonine, there was a wide variation in certain data points (see Figure 3.2.34). On the other hand, the assays that were used to determine the MOI of the time-dependent inhibitors discovered through screening were simpler to carry out, as the enzyme was pre-incubated with the inhibitor for a fixed period of time, before the reaction was initiated with substrates. In this way, the assays were very similar to the MOI assays for rapidly binding inhibitors. However, these assays could not observe the effect of different inhibitor and substrate concentrations on the

rate of inactivation of the enzyme. A more complete picture can be obtained by combining the two assays designed for studying time-dependent inhibitors, but both assays still have the disadvantage of being time-consuming.

An alternative approach to studying time-dependent inhibitors is to follow the time course of a reaction in the presence and absence of inhibitors to measure the effect of time-dependent inhibition. Time-dependent inhibitors will tend to cause the rate of product production over time to become non-linear, as illustrated in Figure 4.1.19.

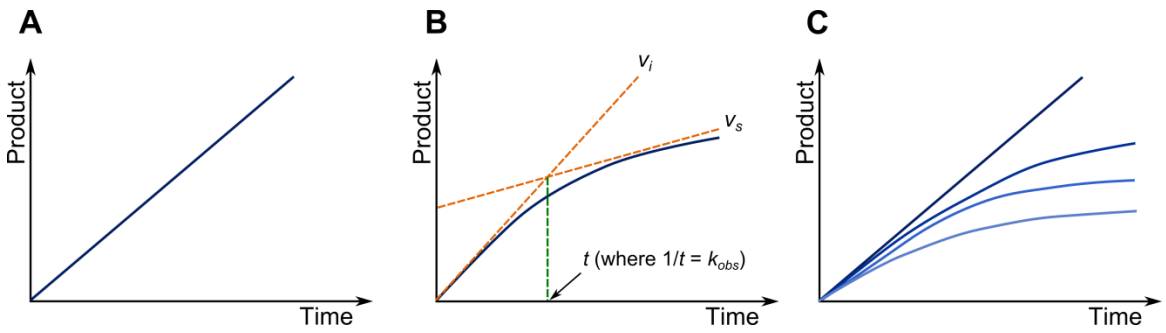


Figure 4.1.19 - the investigation of time-dependent inhibition by time-curve analysis. A = plot showing linear product formation over time by an enzyme in the absence of an inhibitor. B = plot showing a deviation from linearity of the velocity of product formation, *v*, caused by time-dependent inhibition. The initial velocity, v_i, gradually slows to a new steady-state velocity as the time-dependent inhibitor takes effect. C = plot showing the effects of increasing concentrations of a time-dependent inhibitor. The deviation from linearity occurs at an increasingly earlier point in time as the inhibitor concentration and the rate of inactivation, k_{obs} increase. The creation of this illustration was guided by similar illustrations in the literature(194,203).

In the absence of a time-dependent inhibitor, the relationship between product concentration, [P], and time is related linearly with velocity, *v*, and this relationship can be described by the following equation:

$$[P] = vt \qquad \textbf{38}$$

The addition of a time-dependent inhibitor causes the time curve to become non-linear (see Figure 4.1.19, B). As the inhibitor slowly binds, the initial velocity, v_i, is gradually reduced to a slower steady-state velocity, v_s. The rate of enzyme inactivation, k_{obs}, can also be derived from the rate of conversion between v_i and v_s. Fitting equation 39 to measurements of [P] or absorbance over time can be used to derive the v_i, v_s and k_{obs}.

$$[P] = v_s t + \frac{v_i - v_s}{k_{obs}}\left[1 - e^{(-k_{obs}t)}\right] \qquad \textbf{39}$$

This type of analysis would allow k_{obs} to be determined in single sets of replicate assays instead of in multiple assays conducted following different pre-incubation times. For this analysis, the formation of product over time would have to remain linear over the time-course of inhibition. This could be difficult to achieve and would require extensive optimisation.

Further information on the thermodynamics of binding can be obtained from complimentary biochemical techniques such as ITC, microscale thermophoresis (MST) and surface plasmon resonance spectroscopy (SPR), following the attainment of functional data.

4.2 Screening for the discovery of new inhibitors

4.2.1 Utility of Virtual Screening

It is difficult to assess the success of the virtual screening exercise in isolation, as its true utility is proved by its ability to identify *in vitro* inhibitors of the drug target. The hit rate for the 'Custom library', which included compounds identified by virtual screening, was 16.7%. Compared to a hit rate of 0.7% for the Asinex library, and an overall *in vitro* hit rate of 1.4%, it is clear that the *in vitro* hit rate was significantly higher for compounds identified *in silico*. Even excluding compounds from the same chemical series (i.e. discounting myricetin or quercetin and all but one GalE inhibitor), the hit rate is 9.3%. One important feature of the overall screening strategy was that the diversity of the screening library was maintained, whilst virtual screening was used to identify compounds that were more likely to be efficacious *in vitro*. Therefore, amongst the resultant hits there are compounds with diverse chemical structures. Furthermore, the representation of different TDH conformations as the target in virtual screening may have played a role in identifying inhibitors with different binding modalities. Another contributing factor to the effectiveness of the virtual screening experiments was the selection of molecules with known bioactivity for inclusion in the screening library.

Virtual screening has found increasing use in recent decades, both by academic and industrial research groups. By being able to select a subset of molecules that are more likely to be effective *in vitro*, it can save on time and money by allowing one to screen smaller libraries of compounds *in vitro.* There are now several available virtual screening or molecular docking programs. As AutoDock is freely available and the most cited of such programs(29), it was chosen for screening in this study. There was also a wealth of information on the use of AutoDock, due to the fact that it is so widely used. It is difficult to rank the performance of any docking program over others, and the performance of individual programs is said to vary from target to target(29,340,341). Therefore, accuracy of predictions was not a major criterion for selecting AutoDock. Moreover, most docking programs share the same limitations, such as modelling protein flexibility, so those discussed below may be applicable to virtual screening in general, and not just to AutoDock(29).

Most virtual screening or docking programs have been historically limited by their inability to model the conformational flexibility of the drug target. Although programs such as AutoDock4 can model the flexibility of amino acid side chains, modelling larger conformational changes of the target remains a large obstacle to accurately representing drug targets. A common strategy to overcome this limitation is to run a virtual screen multiple times against different conformational states of the target(342–346). Having the privilege of possessing crystallographic structures of different conformations of TDH, it was possible to follow a similar strategy in this study. The significant differences between the hit rates and the highest scoring

compounds in each of the four virtual screens clearly shows that it was necessary to screen against multiple conformations to identify all of the hits that later proved to be effective *in vitro*.

A rigid TDH structure was used, based on AutoDock's ability to correctly predict the binding poses of known ligands to this structure (see Section 3.3.2). This also appears to have been a rational choice, as there is little or negligible variation in the conformations of the side chains of active site residues in the crystallographic structures of TDH obtained. However, with the available knowledge, it was not possible to know if the rigid structures were able to truly predict the binding of molecules larger than L-threonine and glycerol (ligands used to validate the virtual assay). When the various flexible and rigid TDH structures were validated for the prediction of L-threonine binding poses, structures in which Ser82 and/or Tyr144 were flexible, performed similarly to totally rigid structures. It is possible that allowing flexibility for these two important L-threonine-binding residues may have improved the accuracy of the binding predictions for other ligands. Additionally, modelling certain NAD^+-binding residues as flexible may have significantly affected docking predictions, especially in Virtual Screens 3 and 4. By screening only rigid structures, it afforded some consistency to the experiments, such that the changes in results could be attributed mainly to the differences in the overall conformation of the TDH model.

In this study, compounds were selected from the hits of different Virtual Screens, but it was not possible to know whether the results of any one screen were more valid than the other. Biasing the selection of compounds based on their predicted binding energy may have helped to identify compounds that would be more potent binders, but this depends on the assumption that the binding energy predictions are accurate. Unfortunately, the accurate prediction of binding energies is the biggest limitation of molecular docking programs(29). Some researchers have addressed this issue by selecting compounds on the basis of consensus or average docking scores after screening against ensembles of different target conformations(346). This may not have improved results in the current study, as some of the hits were only identified in one of the four screens. However, this approach could become more useful on the availability of more TDH models with different conformations. One additional factor that may affect the efficacy of the chosen TDH models to predict the binding of potential ligands is the absence of water molecules. Water molecules are known to be important in the binding of several ligands, and this has been shown to be the case for L-threonine (see Figure 3.1.39). In a study of ligand docking to cytochrome P450 and thymidine kinase, the consideration of water molecules during docking was shown to improve the predictions of the binding poses of ligands(347). X-ray crystallographic structures of TDH were solved at sufficient resolution to discern the positions of water molecules important for catalytic activity and/or ligand binding. However, water molecules were not considered in these virtual screening experiments due to an early lack of knowledge about the benefits of modelling them. Furthermore, de-solvation effects are already modelled by AutoDock, so the usefulness of modelling water molecules is limited until more information on how a greater range of TDH ligands interact with the protein and water molecules is

obtained. Clearly, these factors can only be taken into consideration in the event of any future virtual screens against TDH or similar targets.

The chemical diversity of both virtual and *in vitro* screening libraries is frequently stated in the literature(348–351) as being important for discovering novel hits. The NCI Diversity Set II provided a convenient source of diverse molecules. The NCI Diversity Set has been used successfully to discover novel ligands for other drug targets(346,351). The ligand substructure library was designed to select compounds that would share the same ligand-protein interactions as the known TDH ligands: glycerol, pyruvate and L-threonine. This approach is proved successful by the identification of the inhibitor 2-{[3-(4-bromophenyl)-3-oxopropyl]amino}-3-hydroxybutanoic acid (BPOB), an L-threonine analogue. However, the Ligand Substructure Library had the lowest hit rate in each Virtual Screen. A particular disadvantage of basing a library on a similarity search of small flexible molecules is that it results in a series of compounds with flexible scaffolds which would suffer higher entropic penalties on binding. The average number of rotatable bonds of compounds in the Ligand Substructure library is much higher than those in the natural product library, despite having a similar average molecular weight (see Table 3.3.1).

The addition of a natural product library was successful in increasing the representation of chemical space within the screening library. The high hit rate of the natural product library may be a result of various factors. The decision to select compounds on the basis of reports of trypanocidal activity of individual compounds or their natural sources may have selected for compounds with activity against TDH(219). However, many natural products have a high number of hydrogen bond donors and acceptors, so they can participate in a high number of interactions, leading to a favourable docking prediction(219). In reality, such molecules are also likely to bind non-specifically to several different sites on a particular molecule and not necessarily to the active site that was the focus of the *in silico* search. It is this character of many natural products that makes them promiscuous ligands(352) that exhibit biological activity by interacting with a range of different targets(216,352). Hence, a particular compound with demonstrated trypanocidal activity may later be found to be acting on a number of different proteins. Any particular natural product has an evolutionary basis behind its existence and it is highly likely to influence some biological function(216,353). Thus, a natural product library can be expected to contain a higher proportion of compounds with biological activity. Although compound libraries such as the NCI diversity sets contain naturally-derived compounds, here it has been demonstrated that the addition of a library built solely with natural products can improve the success of a virtual screen. A similar strategy to that taken here, where a virtual screening library was built by selecting compounds with reported *in vitro* trypanocidal activity, was employed by Ogungbe and Setzer in a screen against three trypanosome targets. In that study, it was found that several *in vitro* trypanocidal compounds were also identified *in silico* as ligands of important trypanosome drug targets. Some of these ligands were predicted to bind

several of the targets included in the study, emphasising the pleiotropic character of some natural products, a property which may or may not be advantageous in drug design(216,218). Promiscuous natural products may possess privileged structures, a core structure that allows them to interact with a broad range of unrelated targets(354). In this respect, identifying such compounds would be desirable as a starting point for lead design(355).

Protein structure similarity clustering (PSSC) was originally proposed to identify molecules to aid the design of screening libraries(216,232) and it has been used successfully in this application(356). PSSC does not appear to have been widely adopted in the identification of novel hits through virtual screening, but this present study serves as a case in support of its usefulness. The majority of the *in vitro* hits identified by virtual screening were identified by PSSC. As the 3D configurations of TDH and the proteins identified by PSSC are so similar (see Figure 3.3.1), inhibitors of those proteins are likely to exploit analogous binding partners. Moreover, as the inhibitors were selected on the basis of having experimental evidence showing that they inhibit the TDH-related target, this probably meant that they were more likely to inhibit TDH than compounds that were randomly selected (e.g. diversity set compounds).

As with the PSSC library, the previously known bioactivity of some molecules, such as the *Mm*TDH inhibitors, probably increased the likelihood that they would be active against TDH. The compounds which have analogous structures to known inhibitors (e.g. School of Pharmacy compounds that are similar to TETD) may also show activity against TDH. In future studies it would be of great interest to see if the activity of other hits of the Knowledge library have comparable efficacy to Qc1 and other inhibitors identified thus far.

One measure of the effectiveness of virtual screening is to compare the predicted binding poses to those determined experimentally. There are no crystal structures of TDH bound to any of the inhibitors identified through screening, but the enzyme inhibition assays provide an insight into the likely binding modes of some of the inhibitors. Judging from the apparent MOIs of the inhibitors identified by AutoDock, it appears that the virtual screening experiments did not predict the binding modes of all ligands accurately (see Section 3.4.4.3). AutoDock seemed to correctly predict that sanguinarine occupies parts of both the L-threonine and NAD^+ binding sites, as sanguinarine was determined *in vitro* to compete with both NAD^+ and L-threonine. On the other hand, BPOB, which was also predicted to bind in the pockets of both substrates, has a noncompetitive MOI towards L-threonine and NAD^+. Qc1, which was predicted by AutoDock to compete with NAD^+, actually appears to be uncompetitive with NAD^+, so preferentially binds the TDH-NAD complex. These two examples reflect an inability of the virtual screening experiments to model the inhibition of these inhibitors – the screen focused in and around the substrate binding sites, but may have excluded yet-to-be-discovered allosteric sites. In other cases, the predictions made by AutoDock may have just been incorrect. Myricetin was predicted to bind in the NAD^+ binding site, but in enzyme assays it exhibited an MOI that was

noncompetitive towards NAD^+. It is possible that the correct binding pose was predicted by AutoDock, but this was not predicted in such a repeatable manner as the seemingly incorrect one (binding poses from the largest cluster of docking results [*n*=50] were considered as the result).

In Section 3.3 it was noted that the virtual screens seemed to select for molecules of higher molecular weight and greater numbers of H-bonding partners. This was likely to be due to the ability to form more extensive interactions with the target which would contribute to a higher binding energy. The issue with such compounds is that they tend to possess a number of unfavourable properties that would make them undesirable drug candidates, as indicated by their non-conformance with rules such as the Lipinski rule of five(277). However, although they serve as useful guides, the rigid use of different rules and metrics in drug discovery has been criticised in recent years, as not all successful drugs adhere to such rules(357,358). When building the virtual screening library, it was decided that compounds should not be filtered for drug- or lead-likeness at such an early stage. Even hits that are unsuitable to serve as lead compounds may provide valuable information, such as important TDH-ligand interactions.

4.2.2 *In vitro* screening design

The selection of compounds for *in vitro* screening was carried out in a similar manner to virtual screening. The compounds identified by UCL ChemiBank using virtual screening began with a library designed for diversity, the Asinex library(244). Screening fragment libraries expanded the approach, so that fragment-based discovery could be pursued alongside a standard lead discovery effort. The fragment based libraries were also designed for structural diversity, so a chemically and structurally diverse range of small molecule and fragment hits were identified. The identification of these hits was enabled by an appropriately-designed assay. The merits and shortfalls of the *in vitro* screening assay are described below.

An advantage intrinsic to the target protein was that the assay for TDH kinetics was relatively simple and therefore, easy to scale up for medium throughput screening. However, some difficulties were encountered on scale-up, the main one being that the enzyme was not sufficiently stable in the reaction buffer for the duration of the time taken to screen an entire microplate. This led to a high false positive rate during initial screening experiments and much time was lost validating inactive compounds. The value of the NIH plate uniformity test (PUT) as a validation tool was highlighted by this situation (see Section 3.4.2). The PUT test indicated that the assay did not meet recommended thresholds for Z'-score and signal window, which are indicators of the ability of an assay to distinguish between active and inactive compounds. However, after the enzyme stability issues were solved with the addition of 1mM potassium

chloride to the reaction buffer, the PUT indicated that the assay was much closer to the recommended state, and accordingly screening with that assay format produced far fewer false positives. As only 1912 compounds were screened in total, the sub-optimal screening assay during the initial experiments did not significantly affect the study. However, in a larger experiment, validation of the assay in the final screening format using a tool such as the PUT is recommended, as the number of false positives could become unmanageable if the assay was not optimised.

Despite the difficulties encountered, using the same general format of the TDH assay in the screen made the scale-up process simpler. A kinetic assay was used, so that artefacts interfering with individual readings would not invalidate an entire assay, as might be the case if a single reading was taken (i.e. in an end-point assay). When testing a wide range of compounds, several of which were coloured, there was the potential for interference with the assay if a test compound absorbed strongly within the region of 340nm. This was not a significant problem at the concentration ranges used to screen the compounds ($\leq$500 µM), so an absorbance assay was used, rather than re-designing the assay to measure fluorescence.

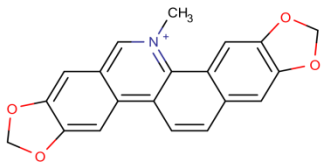
An important choice to be made in designing screening experiments is defining activity thresholds for accepting compounds as hits. The intention is to identify a diverse range of promising hit candidates. However, it is important to set the thresholds low enough so that one has a manageable number of compounds to investigate further. With these considerations in mind, the IC_{50} thresholds of 500µM and 100µM, were optimal in this regard. In fragment-based drug discovery, the original hits are often of a very low affinity, but when linked to other fragments they form a high-affinity compound(20). There is a risk that, by setting an IC_{50} threshold as stringent as 500 µM, several useful fragments would not be identified. However, taking into consideration the comments above, a good number of fragments were identified with the thresholds that were used. Furthermore, there is an opportunity for hit expansion by testing analogues of the fragments. Therefore, it is not a disadvantage that less fragment hits have been identified than may have been identified by setting less stringent activity thresholds.

As has been previously stated, it is not only important to identify chemically diverse hits in a screen, but it is also important to identify mechanistically diverse hits(359). The substrate concentrations used for screening were fixed close to their half-maximal concentrations in an attempt to identify hits that had competitive, noncompetitive and uncompetitive MOIs. Looking at the compounds for which an MOI has been characterised, it is evident that the screen was somewhat successful in achieving this. However, Qc1 is one of the most promising inhibitors identified, but it was not identified using the screening assay format. This compound showed very little or no inhibition in the presence of low concentrations of NAD^+ (1.2mM), but it was actually one of the most potent inhibitors at high NAD^+ concentrations. Thus, there is a case for designing future screens which test the inhibitors under different conditions: L-threonine and

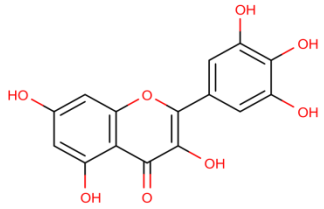
NAD^+ at half-maximal concentrations; L-threonine at half-maximal concentration and NAD^+ at saturating concentration; L-threonine at saturating concentration and NAD^+ at half-maximal concentration. Such an assay design would probably require greater quantities of reagents, but it would ensure that a broader range of hits were identified. Another mechanistic aspect to be taken into consideration is the time-dependent inhibition of TDH. It was necessary to pre-incubate all the test compounds with TDH prior to initiation of the enzyme assay. Several time-dependent inhibitors of TDH were identified with a minimum pre-incubation time of 30 minutes. However, given that the time taken for maximal inhibition of TDH by some time-dependent inhibitors can exceed more than 90 minutes, it can be argued that a pre-incubation time of 60-90 minutes is preferred, to ensure that very slow-binding inhibitors do not become false negatives.

4.2.3 Most promising inhibitors

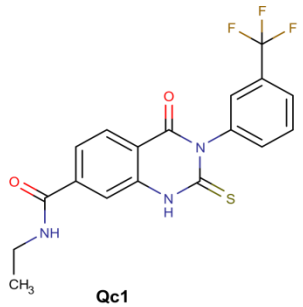
After screening, hits were validated and the MOIs of some hits were investigated in more enzyme assays. Finally, most of the hits were tested for their ability to inhibit the growth of cultured bloodstream form *T. brucei*. A few compounds now stand out as promising for future studies, due to the potency with which they inhibit TDH or the potency with which they inhibit trypanosome growth. The compounds include small molecules from various virtual screening libraries, as well as some fragments. In Figure 4.2.1, the chemical structures of the most promising compounds are displayed alongside information relating to their activities.



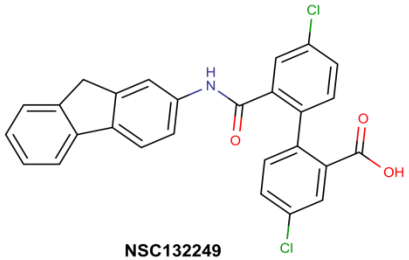
Sanguinarine
Library: Natural Product
VS Hit: VS4
***In vitro* K_i:** 4.5 μM
MOI: competitive (L-threonine & NAD^+)
***T. brucei* growth inh.:** IC_{50} = 0.48 μM



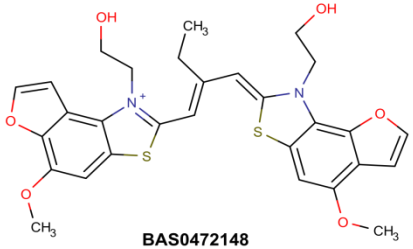
Myricetin
Library: PSSC
VS Hit: VS3, VS4
***In vitro* K_i:** 27.2 μM
MOI: competitive (L-threonine), noncompetitive (NAD^+); TD
***T. brucei* growth inh.:** IC_{50} = 10.9 μM



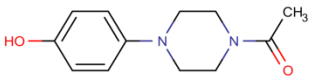
Qc1
Library: Knowledge (*Mm*TDHi)
VS Hit: none*
***In vitro* αK_i:** 25.7 μM
MOI: nonompetitive (L-threonine), uncompetitive (NAD^+)
***T. brucei* growth inh.:** IC_{50} = 7.8 μM



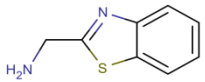
NSC132249
Library: PSSC
VS Hit: VS3, VS4
***In vitro* K_i:** 8.7 μM
MOI: uncompetitive (L-threonine), competitive (NAD^+); TD
***T. brucei* growth inh.:** 25% inhibition at 10.5 μM



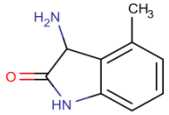
BAS0472148
Library: Asinex
VS Hit: UCL ChemiBank
***In vitro* IC_{50}:** 42.3 μM
MOI: not known
***T. brucei* growth inh.:** 98% inhibition at 8.8 μM



SB02047
Library: Maybridge
VS Hit: n/a
***In vitro* IC_{50}:** 4.6 μM
MOI: not known
***T. brucei* growth inh.:** 60% inhibition at 45.4 μM



CC06013
Library: Maybridge
VS Hit: n/a
***In vitro* IC_{50}:** 132.8 μM
MOI: not known
***T. brucei* growth inh.:** 26% at 49.8 μM



4011502
Library: 3D Fragment Consortium
VS Hit: n/a
***In vitro* IC_{50}:** 111.8 μM
MOI: n/a
***T. brucei* growth inh.:** n/a

Figure 4.2.1 - the most promising TDH inhibitors identified by screening. The structures, IDs, TDH inhibition potencies (IC_{50} or $K_i/\alpha K_i$ [calculated using the Cheng-Prusoff equation]), MOI and *T. b. gambiense* inhibition potencies are displayed.

The hit compounds have different MOIs against TDH, which opens up several avenues for further investigations of TDH inhibition. From a lead identification perspective, it is encouraging to see that all of the compounds adhere fully to the Lipinski rule of five or have just one violation of the rule. All of the fragments in Figure 4.2.1 adhere to the Congreve rule of three. This suggests that they possess at least some of the properties that would allow them to be a starting-point for, or to help guide lead design. At this early stage of drug discovery, it is also important that the hits are amenable to optimisation. This is important during the development of leads or in the determination of structure-activity relationships (SARs). Certainly for the fragments, but also for some of the small molecules, analogous compounds are available that negate the need for the synthesis of new analogues. A few compounds exhibited potent inhibition of trypanosome growth, potentially making them excellent hits for research going forward. The compounds will be evaluated in more detail below.

Sanguinarine

Sanguinarine is a rapid binding inhibitor of TDH that competes with both TDH substrates for binding. Its inhibition of TDH is moderately potent, with a predicted K_i of approximately 4.5 µM. Sanguinarine was originally included in the natural product library because of its documented activity against *T. brucei* and *T. congolense* in culture(360,361).

Sanguinarine was the most potent of all compounds and its IC_{50} for trypanosome growth was 0.48 µM or approximately 0.16 $\mu g.ml^{-1}$. This is within the activity criteria (0.2 $\mu g.ml^{-1}$) previously defined for anti-trypanosome hit discovery(96). A slightly higher ED_{50} of 1.9 µM was reported by Merschjohann and colleagues for the trypanocidal activity of sanguinarine against *T. brucei*(360). Unfortunately, sanguinarine and other alkaloids that are active against trypanosomes are not selective and also tend to be cytotoxic towards human cell lines. An ED_{50} of 1.4 µM has been reported for sanguinarine against human HLA-60 cell lines(360). It could still be of some value to measure how much inhibition of TDH contributes to sanguinarine's efficacy against trypanosomes. However it is highly likely that it interacts with other targets. Most alkaloids are believed to be lethal to trypanosomes by intercalating DNA(360). The most promising use of sanguinarine in the discovery of anti-trypanosome drug targets will be gained by observing the molecular binding interactions that it makes with TDH.

Myricetin (and quercetin)

Like sanguinarine, myricetin was both a moderately potent inhibitor of TDH and a potent inhibitor of BSF trypanosome growth in culture. As a TDH inhibitor, myricetin appears to

compete with L-threonine and with an IC_{50} of approximately 27 µM. The compound is also a time-dependent inhibitor of TDH. The close analogue of myricetin, quercetin shares very similar properties with the inhibitor. Both compounds are inhibitors of dihydroflavonol 4-reductase (DFR) from *Vitis vinifera*(362) and, as DFR is structurally related to TDH, both compounds were identified by PSSC. Quercetin which was also included in the natural product library, is present in the bark and roots of *Securidaca longepedunculata*(363), a plant used in traditional medicinal systems worldwide. Extracts of this plant have been shown to be effective against trypanosomes *in vitro* and they significantly reduced parasitaemia in mice infected with *T. brucei brucei*(364). Over the last decade a number of *in vitro* studies testing myricetin and quercetin have been carried out, and their antitrypanosomal activities have been attributed to various targets(365). For myricetin, these targets include trans-sialidase(366,367) and hexokinase 1(368). Quercetin's activity against trypanosomes has also been attributed to inhibition of hexokinase 1(368,369) and to its metal-chelating properties(370). The affinity of myricetin and quercetin for these other targets (IC_{50} 4-50 µM) (366,369) is comparable to their affinity for TDH. It will be necessary to investigate the contribution of TDH inhibition to the antitrypanosomal activity of these compounds before they can be considered as leads. The fact that both compounds inhibit TDH as well as hexokinase 1, a protein of the glycolytic pathway, is interesting as they may exemplify dual inhibition of the essential pathways for trypanosomal acetyl-CoA production by a single compound.

The myricetin binding mode predicted by docking with AutoDock 4 is not in agreement with that determined using enzyme inhibitory assays. The docking result suggests that myricetin competes with NAD^+ for binding, but the enzymatic data indicates that it actually competes with L-threonine, and not NAD^+. In the case of myricetin and quercetin, and maybe other inhibitors of proteins identified by PSSC, the MOI of the inhibitor may be predicted by its binding mode towards its other target. Crystal structures of DFR bound to myricetin and quercetin show that they bind NAD-bound DFR in the site of the natural substrate(362).

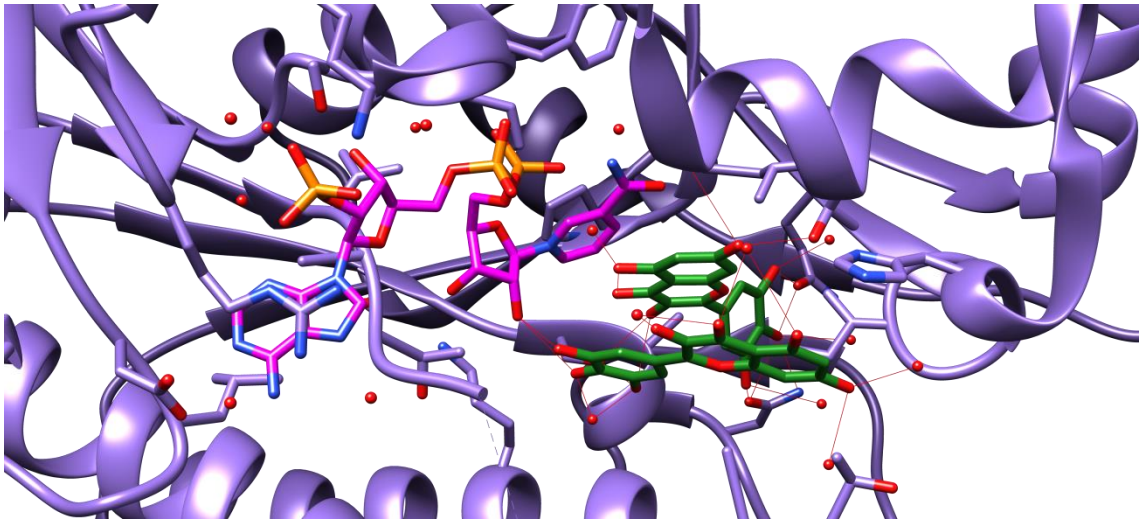


Figure 4.2.2 - binding of myricetin to DFR. Myricetin is coloured green. Two molecules are stacked in the active site.

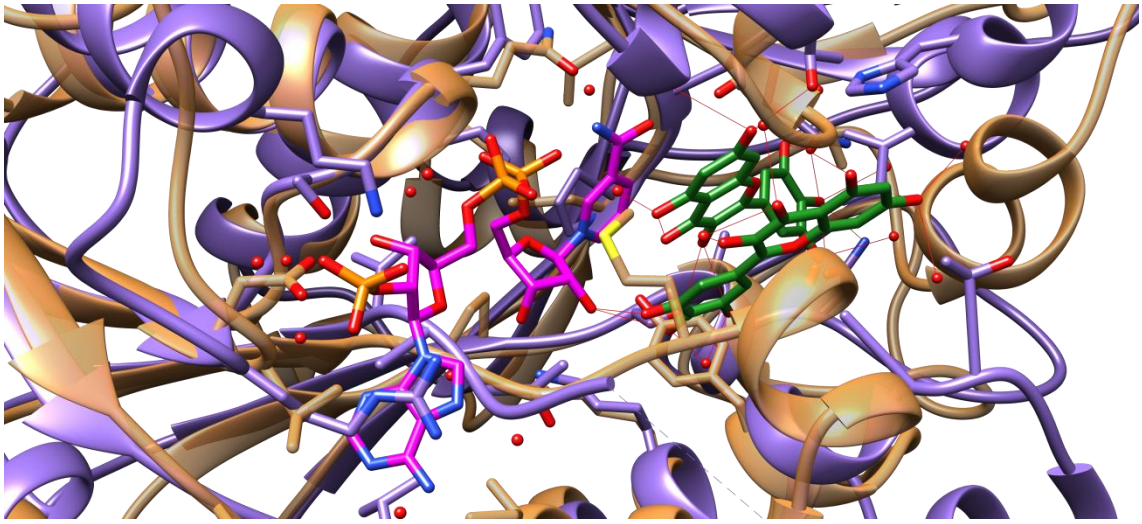


Figure 4.2.3 - TDH structure (semi-transparent beige) superimposed over myricetin-bound DFR. The L-threonine binding site in TDH would not be likely to be able to accommodate two myricetin molecules. However, a single molecule could feasibly bind within the site.

The binding site occupied by myricetin in DFR is analogous to the L-threonine binding site in TDH. The apparent binding mode is unusual, as two myricetin molecules are bound simultaneously in a stacked arrangement. Figure 4.2.3 illustrates that the L-threonine binding site in TDH would not be able to accommodate two molecules of the inhibitor. However it is possible that myricetin could adopt the conformation of one of the molecules shown above when binding to TDH. Quercetin was shown to bind DFR in the same manner, and the same conclusions can be drawn about how that might interact with TDH.

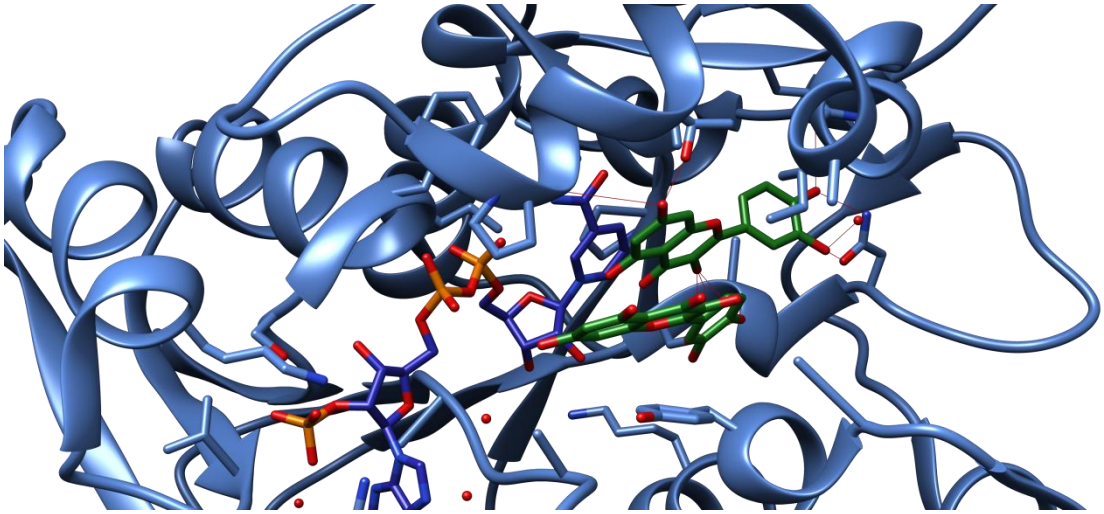


Figure 4.2.4 - binding of quercetin to DFR. Quercetin is coloured green. Two molecules are stacked in the active site, in the same manner as myricetin (see Figure 4.2.2).

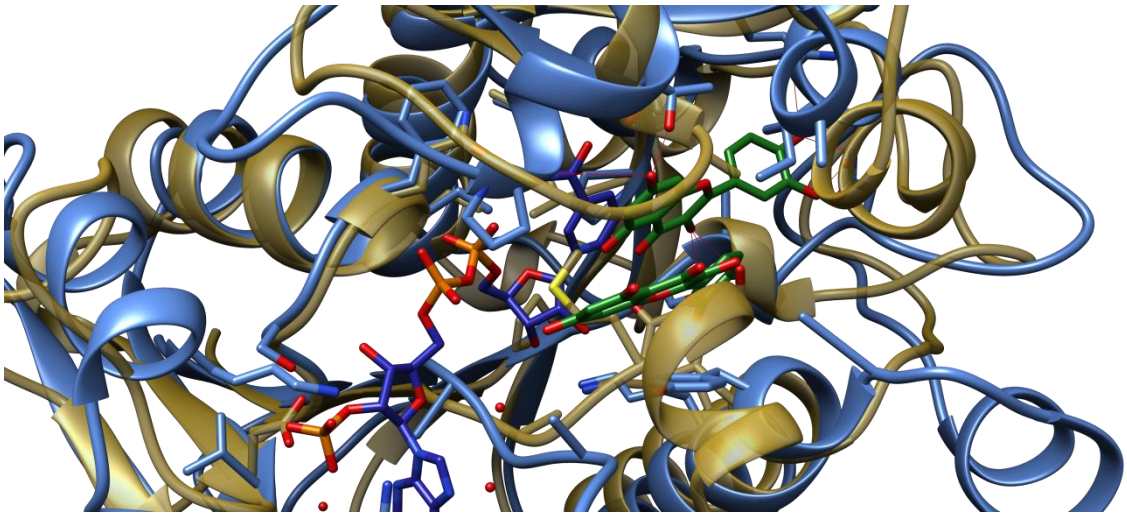


Figure 4.2.5 - TDH structure (semi-transparent gold) superimposed on quercetin-bound DFR structure. As with myricetin, it appears that only one quercetin molecule could be accommodated by the L-threonine binding site.

Figure 4.2.6 shows a crystallographic model of DFR bound to a single molecule of dihydroquercetin(371), aligned and superposed on a TDH structure. It is very similar in structure to the inhibitors myricetin and quercetin, and may serve as an indication of how these molecules bind to TDH. Confirmation of the binding modes of myricetin and quercetin to TDH would be extremely valuable.

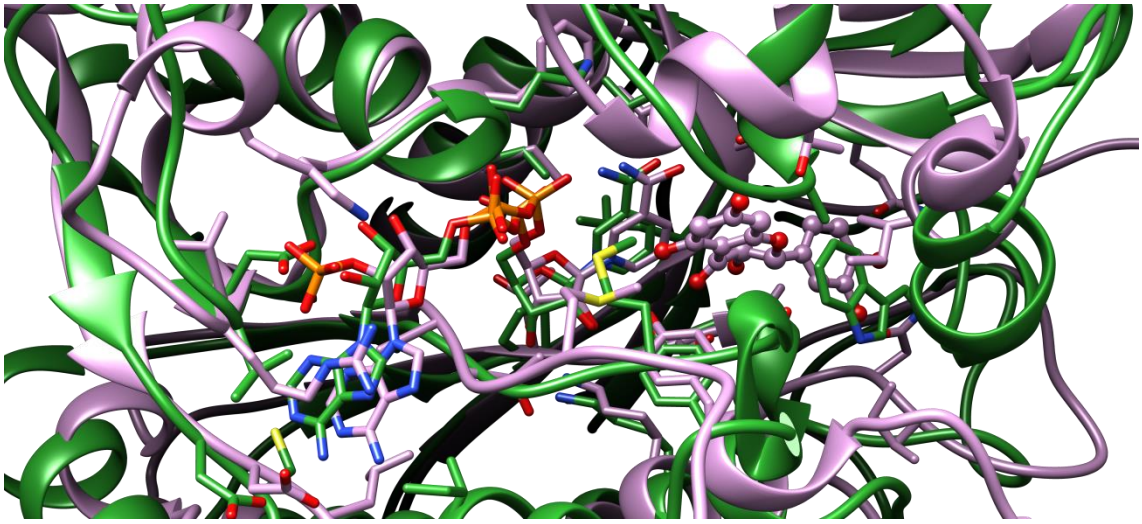


Figure 4.2.6 - TDH structure (green) superposed on a dihydroquercetin-bound DFR structure It appears that the DFR substrate could bind within the L-threonine binding site. Analogues such as myricetin and quercetin could potentially bind TDH in the manner suggested.

Qc1

Under the conditions of the *in vitro* screening assay, Qc1 was not recognised as an inhibitor. This was surprising, given the potent (nanomolar range) inhibition displayed against *Mm*TDH(136), a very similar target. It was discovered that Qc1 is an uncompetitive inhibitor of *Tb*TDH, having much higher affinity for the NAD-bound enzyme. The MOI of Qc1 was noncompetitive to L-threonine. A similar MOI, 'mixed noncompetitive', was reported for Qc1's action against *Mm*TDH, but an uncompetitive MOI towards NAD^+ was not observed(136). Qc1 was one of the most potent inhibitors identified in the presented study, with an αK_i (inhibition constant for binding to the TDH-NAD^+ complex) of approximately 26 µM.

Qc1 was identified as part of a series of quinazoline carboxamide (Qc) compounds in a high-throughput screen against *Mm*TDH. Inhibition of TDH in mouse embryonic stem cells using Qc compounds was lethal(136). It is not known why the activity of Qc1 against *Tb*TDH appears to be several orders of magnitude lower than the activity against *Mm*TDH. It will be interesting to see if the same is observed for other members of the Qc series. Looking at the amino acid sequences of each enzyme (see Figure 3.1.9) there appear to be some differences in the NAD-binding domain that could affect the relationship between TDH, NAD^+ and Qc1.

NSC132249 and *Tb*GalE inhibitors

NSC132249 was a potent time-dependent inhibitor of TDH with a K_i of approximately 9 µM. The MOI towards NAD^+ is competitive and the MOI towards L-threonine is uncompetitive. The binding mode predicted during virtual screen is in some agreement with the observations from enzyme assays: NSC132249 is predicted to bind in the NAD^+ binding site whilst interacting with Loop 1, when it is in the closed conformation. It is possible that NSC132249 binding to the

NAD^+ proceeds L-threonine binding. The closure of Loop 1 following L-threonine binding would allow the inhibitor to bind residues on the Loop, thereby increasing the strength of the interaction. A crystallographic structure would be required to confirm this.

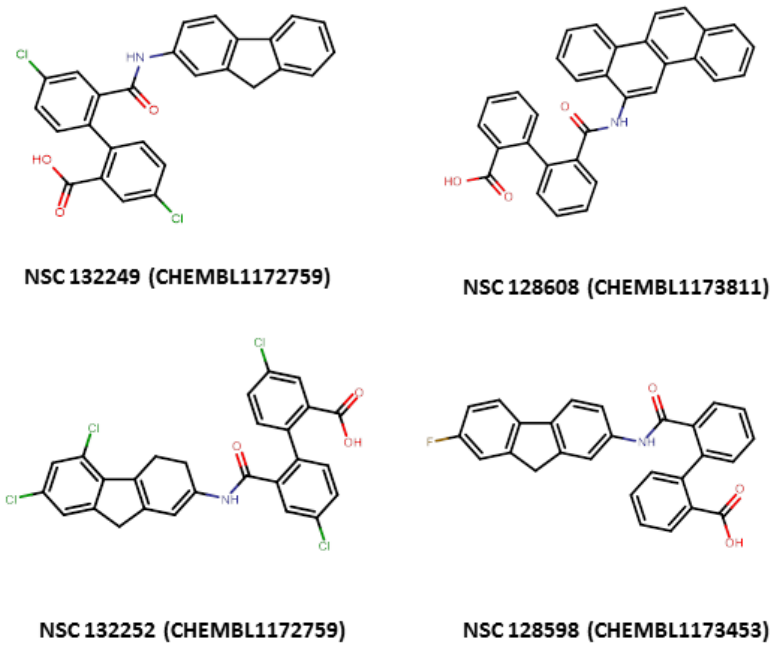


Figure 4.2.7 - inhibitors of *Tb*GalE that also inhibit TDH. Differences in structure may be enough to alter their binding modes and MOIs.

The related *Tb*GalE inhibitor NSC132252 may have a similar binding mode to NSC132249, but the MOI of that inhibitor remains slightly ambiguous due to the different behaviour exhibited in the absence and presence of Triton X-100. NSC132252 may be a promiscuous inhibitor. The differences between the structures of the GalE inhibitors identified as hits may be significant enough that the binding modes and MOIs are also different. For this reason, a single binding MOI and binding mode cannot be assumed to be shared amongst the series without further information. As *Tb*GalE was identified by PSSC, comparison of the structure with that of TDH may provide some clues as to how the GalE inhibitors could interact with both enzymes.

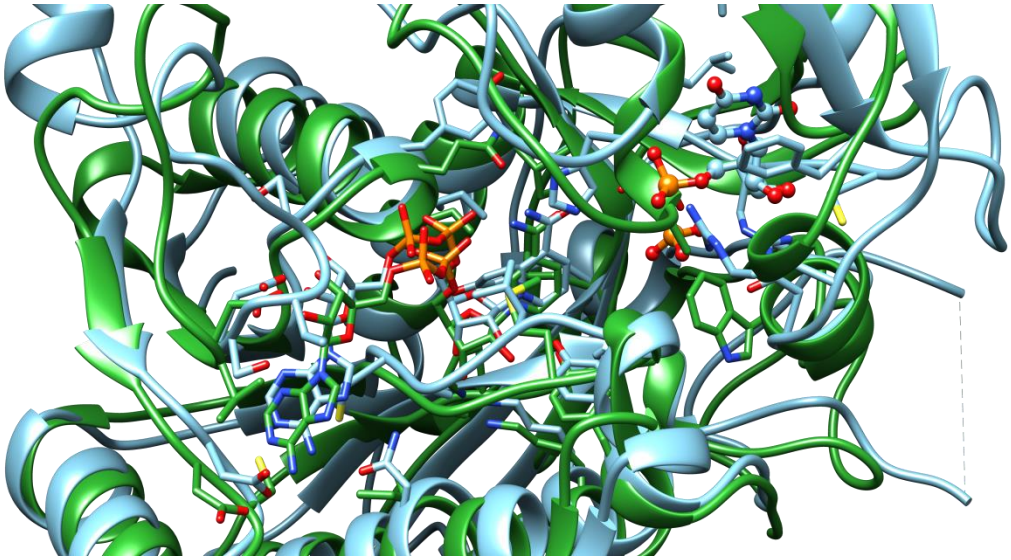


Figure 4.2.8 - TDH structure (coloured green) superposed on *Tb*GalE structure (light blue) bound to UDP (light blue, ball and stick). L-threonine lacks the UDP binding pocket.

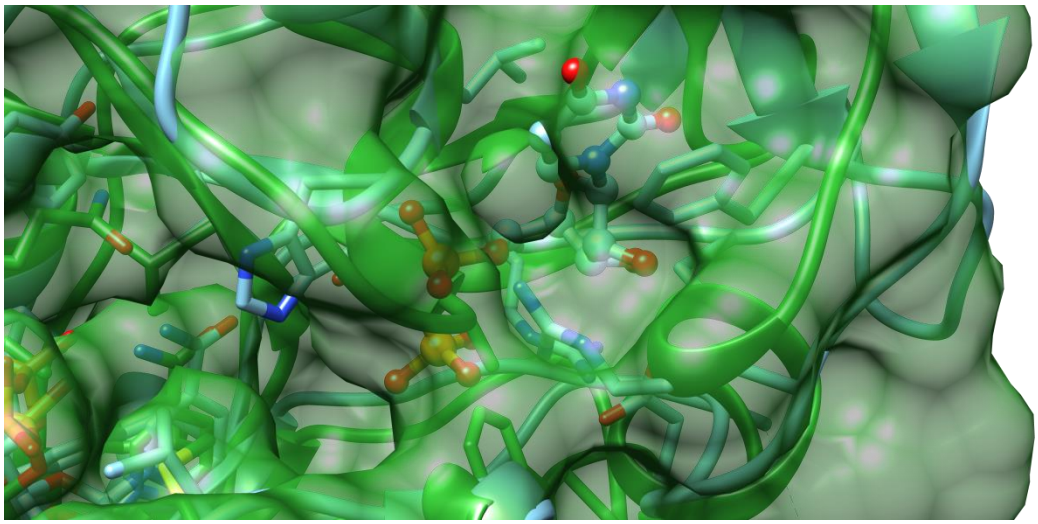


Figure 4.2.9 - TDH structure (ribbon and surface representation) superposed on UDP-bund *Tb*GalE structure. There is no available pocket for UDP to bind to TDH.

The substrate binding sites of TDH and GalE differ markedly, in the respect that TDH lacks a binding pocket, where GalE usually accommodates UDP binding (see Figure 4.2.8 and Figure 4.2.9). The GalE inhibitors were identified in a virtual screen against the target(346). Figure 4.2.10 shows the predicted binding mode of the *Tb*GalE inhibitors. The two arginine residues that are predicted to bind the inhibitors are present in the UDP binding pocket, so, if this virtual screening prediction is correct, the inhibitors must bind TDH by a different modality.

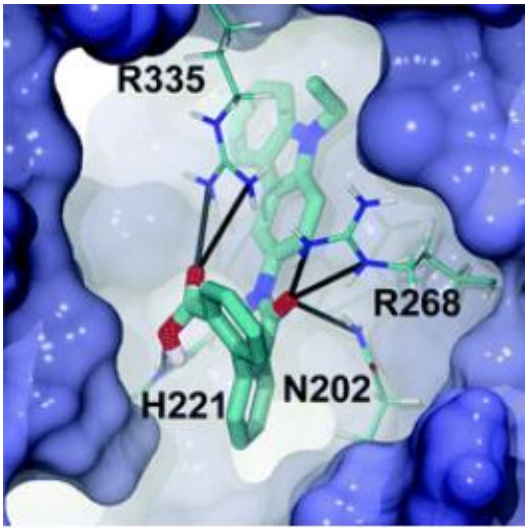


Figure 4.2.10 - binding mode of a NSC132249-related GalE inhibitor to *Tb*GalE, predicted by virtual screening(346).

Although the anti-trypanosomal activity of NSC132249 and the other *Tb*GalE inhibitors is less than that of sanguinarine, myricetin and Qc1, they are of interest because they are reported to inhibit a target that is closely related to TDH. In the same study that identified them as *Tb*GalE inhibitors, they were shown to inhibit the growth of cultured *T. brucei*. NSC132249 demonstrated similar antitrypanosomal efficacy to that presented here and showed an EC_{50} of 28.5 µM. The reported antitrypanosomal activities of other GalE inhibitors are also comparable to those reported here, apart from NSC128598 which showed significant activity in this study, but negligible activity in the *Tb*GalE study(346). It is possible that some of this anti-trypanosomal activity can be attributed to TDH and not GalE alone. This will be important to establish, as the authors of the study hypothesised that the activity of the compounds could be increased by reducing their lipophilicity, thereby increasing cell penetration(346). If the mechanism of action is by dual inhibition of GalE and TDH, then this would also be an important consideration for improving the activity of the compounds. The affinity of the GalE inhibitors for GalE appears to be higher than for TDH. In contrast to TDH, GalE has been shown to be essential in *T. brucei* and *T. cruzi*(372–375), so this is more likely to be the principal target. However, any contribution of TDH inhibition to the antitrypanosomal action should be confirmed.

BAS 0472148

BAS 0472148 is the most promising compound from the Asinex library after it showed potent inhibition of trypanosome growth in preliminary assays. It was unfortunate that the supply of this compound was limited, as its activity is comparable to the other promising compounds mentioned above. Little information appears to have been published on this compound. In addition to the accurate determination of its inhibition of trypanosome growth, the contribution of BAS 0472148 inhibition of TDH should be investigated as a priority. BAS 0472148 is also quite complex, so the binding mode should be investigated by X-ray crystallography.

Fragments

There are up to eight fragments identified in this assay that warrant further study. Of these fragments, two compounds from the Maybridge library, SB02047 and CC06013, and one compound from the 3D fragment consortium library, 4011502, are particularly promising due to their moderate to high affinity for TDH. The Maybridge compound SB02047, in particular, demonstrated very high potency for a fragment, and this potency is similar to that of the larger inhibitors. As one might expect, none of the fragments showed potent inhibition of trypanosome growth in culture. Despite this, the current hits serve as a good starting point for fragment-based drug design. The binding interactions of the most potent fragments may also give an insight into a structural feature of TDH that can be exploited for increasing the potency of all candidates.

Other compounds

If time or resources persist, much valuable information can be gained by investigating some of the other hits, as discussed for the most promising compounds. The enantiomeric purity of BPOB was not specified by the manufacturer/wholesaler, and there is the possibility that only one stereoisomer is active as a TDH inhibitor. A pure form of the active stereoisomer would therefore exhibit higher potency than that exhibited by the BPOB tested so far. The Asinex compounds collectively represent a diverse range of structures, so determining the interactions made between these compounds and TDH would be very valuable for future lead design.

4.2.4 Lessons for the design of screening campaigns

The strategy taken in the design of the virtual and *in vitro* screening libraries can be considered one of the main reasons for the success of those experiments. A focus on achieving diversity has resulted in a number of chemically and mechanistically diverse inhibitors of TDH. It is also evident that taking a knowledge-based approach, which involved screening ligands with known bioactivity, particularly against trypanosomes and enzymes structurally related to TDH, enriched the library and increased the number of hits. It is important to employ a balance of empirical and knowledge-based strategies, however. Although there is evidence to suggest that compounds identified through virtual screening were more likely to be *in vitro* hits, a screen relying solely on such compounds could preclude the discovery of inhibitors that were not correctly identified by the virtual screening experiments. In this study, the virtual screens carried out with AutoDock 4 and with GOLD (carried out by UCL ChemiBank) focused on the substrate binding sites, so they may not have identified compounds binding to allosteric sites or other sites, such as the dimerisation interface. The screening of fragment compounds not only enabled the pursuit of a fragment-based approach, but also enabled the identification of hits

from a pool of compounds that were free of any biases imposed by virtual screening experiments.

Prior experience with the TDH assay reduced the time to design and set-up the *in vitro* screen. In addition, the knowledge of the different modes of inhibition exhibited by the known TDH inhibitors helped in the selection of substrate concentrations that would allow the identification of diverse hits. As mentioned previously, full validation of the assay in a format similar to that used for screening is highly recommended to ensure accuracy and reproducibility of results. Statistical validation of the assay, using tools such as the NIH PUT assay, is also effective in ensuring the appropriateness of the assay. A well validated assay can save on time and resources, and this is particularly advantageous during a large screen. In a biochemical screen such as the one presented in this thesis, one would ideally want to carry out the following within a reasonably short space of time:

- production and purification of the target protein;
- execution of the screen using the primary assay;
- confirmation of hits using a secondary assay.

This is mainly for reasons of stability – confounding results can be avoided by ensuring that the same batch of protein is used to assay all compounds, that the protein remains pure and intact, and that the compounds do not break down over time. Furthermore, once hit compounds have been obtained from a new source, validation studies should also be carried out across a reasonably short time span for the same reasons.

If possible, hit compounds should be validated by several methods. In this study, where material was available, compounds were tested in enzyme assays to confirm dose-dependent inhibition, determine their MOIs and to test for the disruption of non-specific inhibition using Triton X-100. The compounds will be validated further if crystallographic data is obtained.

4.3 Implications of the Research

4.3.1 TDH as a target for trypanosomiasis

TDH was first highlighted as a drug target for trypanosomiasis by the research of Cross and colleagues in the 1970's. The compound TETD, a human aldehyde dehydrogenase inhibitor marketed as a drug for the treatment of alcohol dependence (disulfiram or Antabuse), was found to potently inhibit TDH. Exposure of cultured trypanosomes to TETD caused their rapid death and this trypanocidal effect was shown to be accompanied by a drastic reduction in the production of acetate from L-threonine(138,139). In the last few years, research has been published on the essentiality of acetate production in both procyclic and bloodstream form trypanosomes. Dual blockade of the two pathways for acetate production, the TDH pathway and the glycolytic pathway, is lethal to trypanosomes(143,144). This effect highlights the therapeutic potential of inhibiting this pathway, which is believed to play a role in several processes, including fatty acid synthesis and energy metabolism(138,141). In other organisms, TDH has also been shown to play a key role in metabolic regulation, allowing organisms to survive in conditions with limited resources. Interestingly, in recent years the TDH pathway has been said to play a role in genetic regulation in mouse embryonic stem cells(137,376). Considering the roles played by this enzyme in various organisms, it is likely that the enzyme has evolved as a key metabolic regulatory enzyme. The TDH gene is upregulated as the morphology of *T. brucei* adapts to new environments in the gut of the tsetse fly(138). Similar changes in the morphology of the parasite are reported to take place within humans as the parasite switches to a transmissible 'stumpy' form(81). Therefore, disruption of the TDH pathway could also be a strategy to limit the spread of the parasite. Looking at all the potential processes that could be affected by TDH, one hopes that new information will become available on other deleterious phenotypes caused by TDH inhibition.

In consideration of the target product profile (TPP) for an antitrypanosomal drug, a key strength of TDH as a drug target is that humans do not have a functional copy of the enzyme, and so metabolise L-threonine using other pathways(106). This presents an opportunity for the discovery of a drug with a lower chance of causing side-effects. One note of caution is that TDH is a member of a very large superfamily of enzymes, many of which are utilised by humans(283). Therefore, in order to achieve the desired TPP, selectivity profiling assays and judicious ligand design to avoid off-target effects will be essential.

In light of the evidence that TDH knockdown is not lethal to *T. brucei*, it has been somewhat surprising to discover so many inhibitors of TDH that inhibit trypanosome growth. As described previously, some of the most promising hits are known to inhibit other drug targets. Previous work by Professor Jon Cooper at UCL and Professor John Kelly at the London School of Hygiene and Tropical Medicine showed that suppression of the TDH gene by gene knockdown rendered cultured *T. brucei* more susceptible to TETD, but this only showed that TDH inhibition contributes partly to the lethality of TETD towards trypanosomes. TETD is known to be an

effective trypanocide towards *T. cruzi*. This effect has been attributed to the compound's metal-chelating ability(377–379) and to inhibition of delta-1-pyrroline-5-carboxylate dehydrogenase(380,381). It is possible that TDH forms an important part of the activities of all of the multi-target compounds discussed in this chapter. If that is true, it is likely that a multi-target drug discovery strategy is the best course of action for developing a TDH inhibitor-based antitrypanosomal drug.

4.3.2 Combination therapy

Whilst TDH blockade alone has not been demonstrated to be lethal to trypanosomes, dual blockade of TDH and pyruvate dehydrogenase, the other enzyme producing acetyl-CoA in trypanosomes, was shown to have a profound effect on trypanosome growth(144). Therefore, there is a strong case for the development of a TDH-inhibiting drug, as part of a combination with an inhibitor of the glycolysis pathway. Compared to a single drug, combination therapies can have numerous advantages, including:

- possibility to achieve greater efficacy than a single drug targeted to a single target;
- targeting more than one target can mitigate the effects of mutation-mediated drug resistance;
- lower dose requirements for both drugs if the effects are synergistic or additive.

Contrastingly, there are several valid arguments against the use of combination therapies in place of single drugs, some of which include:

- if the drugs have different pharmacokinetics, their co-location at the desired site of action may not be easy to achieve;
- side-effects may become more likely as the number of drugs is increased;
- the risk of drug-drug interactions (for patients taking other medication) is increased with combination therapies(382).

In order to truly exploit the benefits of a combination therapy, the effects of the two drugs should be synergistic. For clarity, the concept of synergism warrants a brief explanation. As drug combinations have become more common, particularly in anti-cancer and anti-infective drug design, methods have been developed for the detection and classification of drug effects in combination(383,384). When the effect of two drugs in combination is the sum of the effects of both drugs in isolation, the combination is said to be 'additive'. When the effect amounts to less than the sum of the two drugs' effects in isolation, then the combination is said to be 'antagonistic'. The preferred situation for a drug combination is when the magnitude of the combined effect is greater than the sum of the drugs' effects in isolation, in which case the combination is said to be 'synergistic'(383). The isobologram, a graphical depiction of the effects of drug combinations, shown in Figure 4.3.1 illustrates this concept.

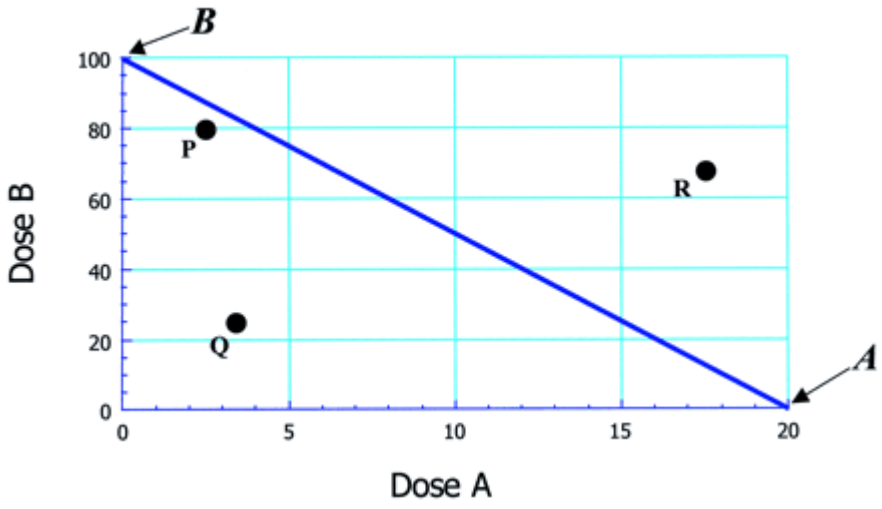


Figure 4.3.1 - illustrative example of an isobologram, published by Tallarida(383). The diagonal blue line represents the combination of doses of drug A and drug B that will achieve the same effect. It is the additive effect of each drug in isolation. A drug combination represented by point P lies close to the line, so can be said to be additive. Conversely, the drug combination at point Q lies below the line, indicating that it is antagonistic. The combination represented by point R is synergistic as it is situated above the blue line.

If a combination of a TDH-inhibitor and a glycolysis inhibitor was to show synergy, then this could prove to be a highly effective approach to the treatment of trypanosomiasis. The magnitude of effect required by this combination would presumably need to be trypanosome death. A synergistic combination, combined in a single dosage form, could help to meet the TPP by:

- reducing the risk of side-effects by requiring lower doses of each drug;
- increasing the activity against drug-resistant forms;
- shortening the length of treatment and simplifying dosage regimens by having a more effective treatment.

The potential for a synergistic combination of a TDH inhibitor with a glycolysis inhibitor should be investigated by use of small interfering RNA or selective inhibitors of each pathway. Combinations of TDH inhibitors with inhibitors of other essential processes could also potentially be explored.

4.3.3 Targeted polypharmacology

Some of the aforementioned benefits afforded by the use of combination therapies can also be achieved by drugs targeting multiple ligands. It seems likely that several of the TDH inhibitors exhibiting *T. brucei* growth inhibition are active at other targets in addition to TDH. Myricetin and NSC132252 are good examples of this. As exemplified by the GalE inhibitors, the inhibitors identified by PSSC may be inhibiting trypanosomes by inhibiting both TDH and the trypanosomal PSSC-identified target. All of these observations highlight the now well-known fact that several pharmacologically active substances mediate their effects through interactions at multiple targets. The rational design of multi-targeted drugs, often called designed multiple

ligands (DML) has been increasingly used in drug discovery(382,385–387). The rational design of a DML targeting TDH and other targets would be greatly aided if the target binding sites either share some structural similarity or a similar arrangement of pharmacophores. For this reason, DMLs are usually targeted at proteins within the same family. Kinases are a good example of such a family of proteins, and there are several examples of kinase inhibitors in development and in clinical use that exhibit serendipitous or targeted polypharmacology(382,388,389). This requirement for target similarity is already met in the case of TDH and GalE. Compounds such as myricetin and quercetin appear to show polypharmacology towards diverse targets. This may be a reflection of the promiscuity of those compounds, but it could also be an indicator of the potential for the design of ligands targeting TDH and another diverse target. Irrespective of this, the rational design of DMLs targeting proteins from different families is understandably much more difficult than designing DMLs for related protein targets(382,385).

Although the discovery of multi-targeted ligands was not an aim of the presented study, a foundation for such an effort has been set. Morphy and Rankovic describe two approaches to the design of DMLs: a “knowledge-based” approach, where pre-existing data on the biological activity of compounds is utilised, and a “screening-based” approach(385). In the present study, both approaches were combined in the hope of increasing the number of hits. As a result, compounds that are active at other targets in addition to TDH have been discovered.

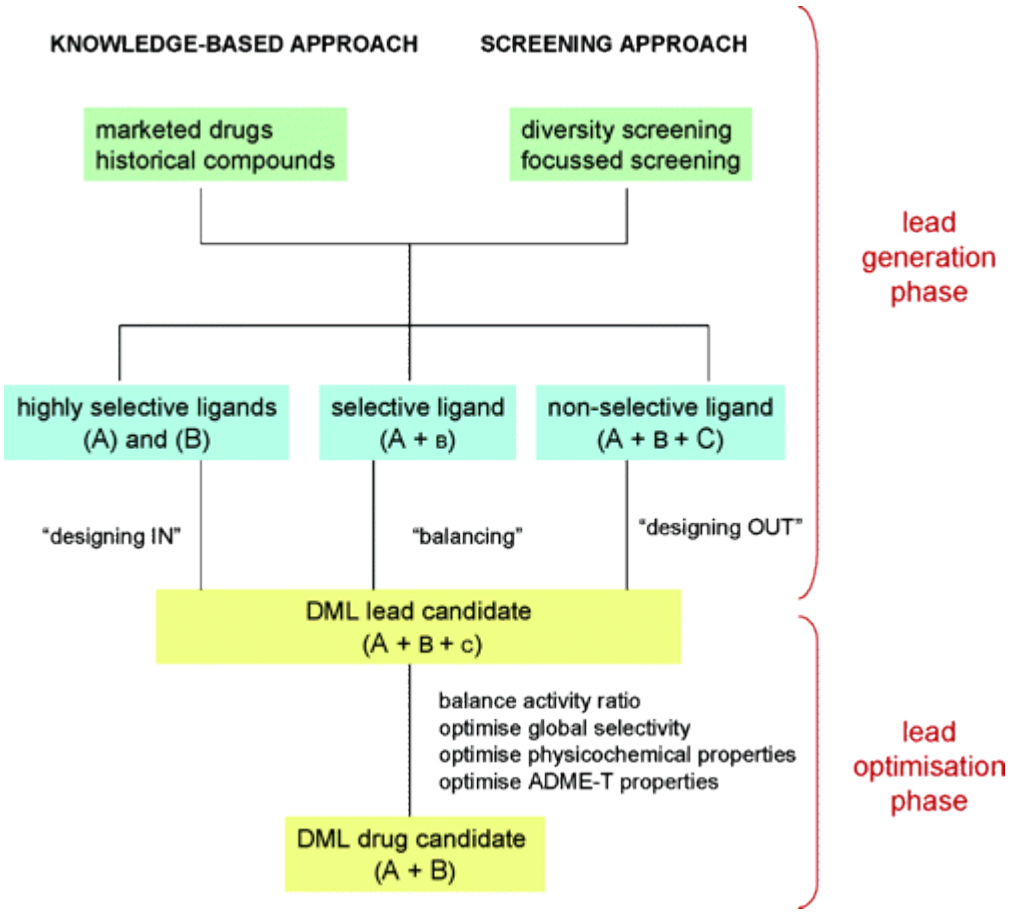


Figure 4.3.2 - strategies for the design of DMLs described by Morphy and Rankovic(385). The earliest phases of the lead generation phase have been conducted in this study. The next stages can begin with selective ligands of the targets, or even non-selective ligands. The design strategy is different depending on the types of ligands available.

Figure 4.3.2 gives a succinct summary of the process that may be followed for the design of DMLs. If this was to be considered an appropriate strategy for antitrypanosomal drug design, then information on the activity (including potency and MOI) of TDH inhibitors towards other drug targets would be required.

4.4 Future research

4.4.1 Further structural and functional studies of TDH

The findings presented here will allow further drug discovery efforts to proceed – a good number of inhibitors have been identified and validated assays are available for the discovery and characterisation of new inhibitors. However, there are still several basic research questions that remain to be answered, and other questions have arisen following the findings presented herein.

Several lines of evidence support the variable nature of the TDH conformation. The data from various structural biology studies suggest that some significant conformational states of TDH

remain to be seen. In particular, the possible dissociation of TDH into active monomers needs to be verified and investigated further to establish if the phenomenon truly occurs. It may be possible to induce dissociation by treatment with *p*-chloromercuribenzoate as was done to induce dissociation of GalE from *K. fragilis*(309). There is plenty of scope for further studies using the techniques used in this study, as well as with other techniques not used here, such as DLiS.

Novel discoveries have been made on the kinetic behaviour of TDH and the inhibition of TDH has been studied in more detail here than the majority of studies of the enzyme. Nevertheless, more information is required to be able to model the action of the enzyme more completely. Firstly, the extent and nature of the substrate/product inhibition can be investigated and its impact on other kinetic parameters determined. Secondly, a deeper insight into the variables influencing the biphasic kinetics of TDH will also help to model the kinetics. From a drug discovery perspective this information will be useful for understanding the MOIs of inhibitors, which are likely to be more complex than standard models of MOIs suggest. Gaining a crystallographic confirmation of the binding modes of various inhibitors could match certain inhibition patterns to different binding sites. This, in turn, would provide more information on the structural determinants of TDH function.

4.4.2 Towards the design of new leads for trypanosomiasis

Without knowledge of the *in vivo* effectiveness of TDH inhibition for treating trypanosomiasis, it is not possible to make a conclusive statement on its attractiveness as a drug target. However, taking several factors into consideration, there are several reasons why pursuing this target may be an effective approach. Frearson et al. published a number of criteria for the assessment of drug target candidates for antiparasitic drug discovery(32). The assessment is reproduced in Table 4.4.1, with an indication of how TDH would be assessed.

Table 4.4.1 - the assessment of drug targets for antiparasitic drug discovery. The table is reproduced from the table presented by Frearson et al.(32). The criteria for assessment are divided into a traffic light system, with 'green' representing favourable characteristics and 'red' representing unfavourable characteristics. The cells representing TDH characteristics are coloured. *Although there is no essential TDH homologue in humans, there is the structurally similar UDP-GalE.

Criterion	Red	Amber	Green
Target validation	No or weak evidence that the target is essential for growth or survival	Either genetic or chemical evidence that target is essential for growth or survival	Genetic and chemical evidence that target is essential for growth or survival
Druggability	No drug-like inhibitors are known and active site of target is not druggable	Drug-like inhibitors are known or active site is druggable potentially	Drug-like inhibitors are known and druggable active site (i.e. clinical precedent within the target family)
Assay feasibility	No in vitro assay developed and/or significant problems with reagents (cost or supply)	In vitro assay exists, development into plate format feasible but not achieved	Assay ready in plate format and protein supply assured within appropriate timelines
Toxicity	Human homologue present and little or no structural or chemical evidence that selective inhibition is possible	Human homologue present but some structural or chemical evidence that selective inhibition is possible	No human homologue present or human homologue known to be non-essential*
Resistance potential	Target has multiple gene copies or isoforms within the same species and is subject to escape from inhibition	Target has isoforms within the same species or might be subject to escape from inhibition	Target has no known isoforms within the same species and is not subject to escape from inhibition
Structural information	No structure of target or closely related homologue	Structure without ligand available and/or poor resolution (>2.3 Å) or opportunity to build a good homology model	Ligand-bound structure of target (or ligand in closely related homologue) available at high resolution (<2.3 Å)

Although TDH does not fare well on what is probably the most crucial part of the assessment, there are good opportunities to simultaneously target TDH and other targets, as discussed above.

There are several approaches that could be taken for the design of a TDH inhibitor. Whilst they may eventually provide valuable information on SARs, none of the current inhibitors stand out as a good candidate on which to base lead design. The current hits are generally quite large, and some of them do not leave much room for optimisation. A potent compound, of medium molecular weight and with a scaffold amenable to optimisation would be ideal. To find such a

compound, and preferably one with higher affinity, the first action to be taken could be to run a larger scale, high throughput screen against TDH. The screening library could be focused using virtual screening and the knowledge gained from the recently conducted screen.

Alternatively, one could choose to begin lead design based on the currently available information and on information gained from the current pool of hit compounds. Once structural data has been obtained on the binding of small molecule hits to TDH, *de novo* drug design could be guided by knowledge of the 3D configurations of important pharmacophores. New scaffolds can be designed as bioisosteres of bound conformations of active compounds(390). A fragment based approach can also be followed. The binding sites of individual fragments can be determined crystallographically, and ligands binding to different sites can be fused to form a compound of higher affinity. As mentioned previously, more potent compounds may be found to replace each fragment by testing commercially available or synthesised fragments. This approach has been exploited successfully to discover potent lead candidates in various fragment-based drug design efforts(20).

Equally, there is scope for rapid SAR development using the small molecule inhibitors. Other members from the Qc series of *Mm*TDH are commercially available. Other *Tb*GalE inhibitors identified by Durrant et al. and potentially many more analogues can be obtained free of charge from the NCI/NIH Developmental Therapeutics Program. Perhaps the best opportunity for SAR determination and the identification of more potent hits lies with the flavonoids myricetin and quercetin. Flavonoids are abundant in nature and widely studied for their bioactivity(365,366), so analogues of myricetin and quercetin could be obtained with ease.

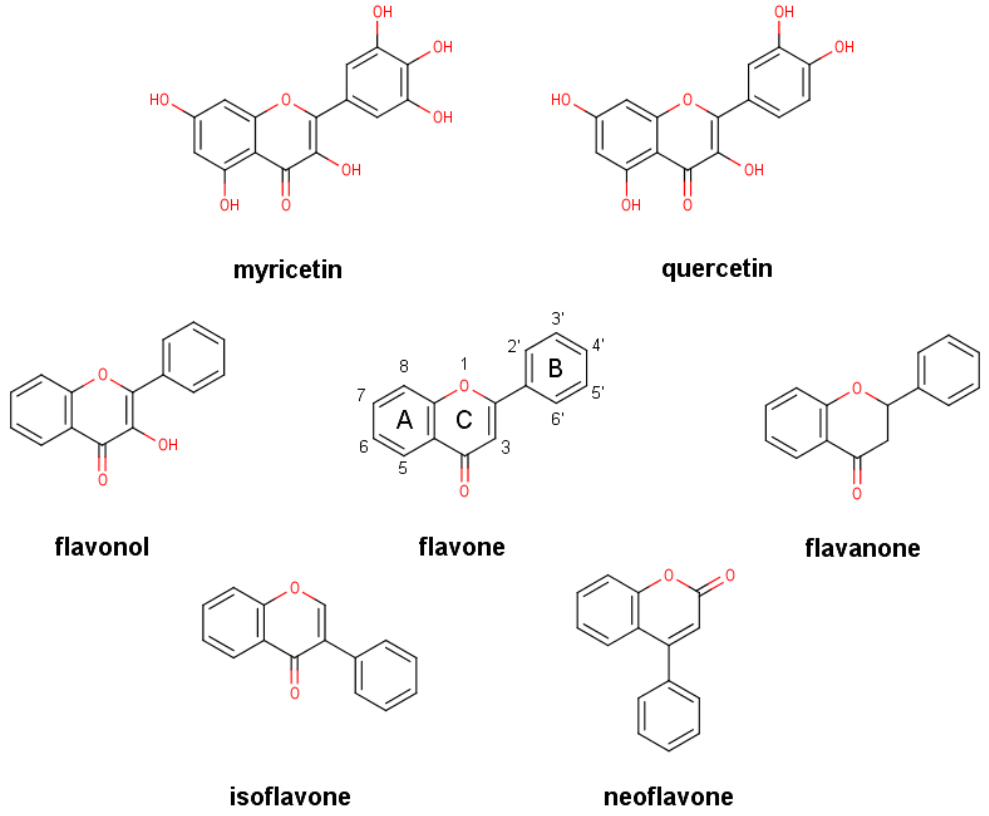


Figure 4.4.1 - myricetin, quercetin and other flavonoid classes. The flavone structure is numbered to highlight all the positions where substitutions can be compared to develop SARs.

Virtual screening results suggested that the most potent compounds are those that can bind within the L-threonine and NAD binding pockets simultaneously. However, at present it is unclear as to whether compounds of a certain size can feasibly do this, due to the location of the side chain of the residue Met81. Different arguments can be made in support of designing ligands that bind exclusively to the L-threonine binding site and those that bind exclusively to the NAD binding site. A molecule binding to the L-threonine site is more likely to be specific, as opposed to a molecule with affinity for the NAD-binding site, which possesses structural features that are widely conserved. Furthermore, as L-threonine binding is mediated by fewer residues than NAD-binding, any mutation that would confer resistance to ligands of this site would be more likely to lead to a loss of function of TDH. However, the L-threonine binding site is small, meaning that it may be difficult for a ligand to form enough strong interactions within such a small area to confer high affinity binding. The size of this site is also influenced by the conformation of Loop 1, which may or may not permit binding of large ligands. The NAD binding site, on the other hand, presents a larger area for a ligand to make interactions with a protein. Therefore, more potent ligands may be designed by focusing on this site. Although structural features of the NAD binding site are widely conserved, it may still be possible to achieve specificity, as demonstrated by the numerous selective kinase inhibitors targeting the ADP binding site(391). Ultimately, the decision on which site to target will be influenced by the

results of X-ray crystallographic studies of TDH and the hit compounds. Studies involving other trypanosomal protein targets may also help to determine if a targeted polypharmacological strategy is feasible.

During the design of new TDH inhibitors, close attention must be paid to the structure-function relationship of the enzyme. The exact nature of ligand-induced conformational changes may dictate which molecules will bind TDH with high affinity. For instance, if NAD and L-threonine are bound by a classical induced fit mechanism, it may be that the most effective inhibitors will be those that have improved affinity upon induced fit binding to the enzyme. Conversely, if the way that TDH binds substrates is more akin to the process described as conformational sampling, where multiple conformations are present in a given population of enzymes, then TDH could potentially bind a more diverse range of inhibitors, as sampled conformational states are only locked once certain binding moieties are present.

The production of acetyl-CoA and glycine from L-threonine relies on the coupled actions of TDH and KBL. Inhibition of TDH alone is likely to be sufficient to inhibit the pathway, as has been demonstrated in *T. brucei* and in mouse embryonic stem cells(135,136,138,139). However, this should be confirmed for each inhibitor, either in coupled assays with KBL, or in cultured trypanosomes. There is a possibility that interaction between TDH and KBL may alter the conformation of the former enzyme, potentially altering the affinity of an inhibitor for TDH.

4.4.3 Targeting other organisms

TDH has been identified in several other organisms of medical interest, including important pathogens such as *Clostridium difficile, S. aureus* and *Enterococcus* spp. Illnesses caused by these pathogens all represent significant health burdens, and there is a need for novel antibiotics to treat multi-drug resistant forms, and to halt the spread of further resistance. Looking beyond systemic anti-infectives, TDH could also be a target for the discovery of new insecticides. The insect vectors of human African trypanosomiasis, Dengue fever, malaria and lymphatic filariasis all appear to have TDH genes. The role of TDH in many organisms remains to be determined. However, there may be opportunities for TDH inhibition in combination with other strategies in each case.

4.4.4 Drivers of NTD R&D

The main drivers of NTD product R&D come from philanthropic organisations, the pharmaceutical industry, governments, PDP brokers and other research institutions (academic and PSNRI). Prior experience in NTD product R&D strongly indicates that partnering with one or several of these organisations would be imperative. The research in this thesis focuses on NTD development from the point-of-view of researchers from an academic institution/PSNRI. Hence, the following discussions will analyse different commercialisation strategies and potential partners from that viewpoint.

Groups from Universities or PSNRIs play a very important role in the discovery of new innovative treatments(69,392), but they also have several limitations as projects progress towards the market(70,393,394). Therefore, partnerships are essential to them.

4.4.5 Commercialisation strategies

4.4.5.1 General strategies for pharmaceutical products

The first step in commercialising scientific research typically entails the generation and protection of intellectual property (IP) pertaining to an invention. With pharmaceuticals, this will usually involve filing a patent. The holder of the patent can then seek to capitalise on this IP through licensing of the invention to a third party, who may have more resources to develop and distribute a marketable product. Another strategy to reap rewards from IP is to form a spin-out company. Funds to develop the invention are raised in the form of equity in the company. The company may enter into a partnership with third parties to share risk and to co-develop the product. The university spin-out company is increasingly becoming a method favoured by Universities to exploit new IP and this plays an important role in new product discovery and development(395). This situation holds true in new drug and vaccine development by Universities(70) and companies alike. Both academic institutions and new companies have been identified as important sources of new drugs(69,392).

Given the high risks and capital costs involved in pharmaceutical R&D, the formation of co-development partnerships is almost ubiquitous. Large pharmaceutical corporations have looked to academic research institutions and small/medium biotechnology or pharmaceutical enterprises (SMEs) as sources of new drugs and vaccines(71). Meanwhile, their smaller, less resourced partners are able to benefit from the extensive capabilities and product development experience of the larger firms. Certain R&D activities are often outsourced to contract research organisations (CROs), often as a cost saving measure, but sometimes as a faster and more efficient way to advance products through clinical development. When new product development is initiated by small companies, and sometimes when large companies are the

sole developer, it may be necessary to bring in a number of other partners (venture capitalists and other shareholders) to finance development.

The following section discusses how different development partners may be involved in the development of drugs for a neglected tropical disease, such as trypanosomiasis.

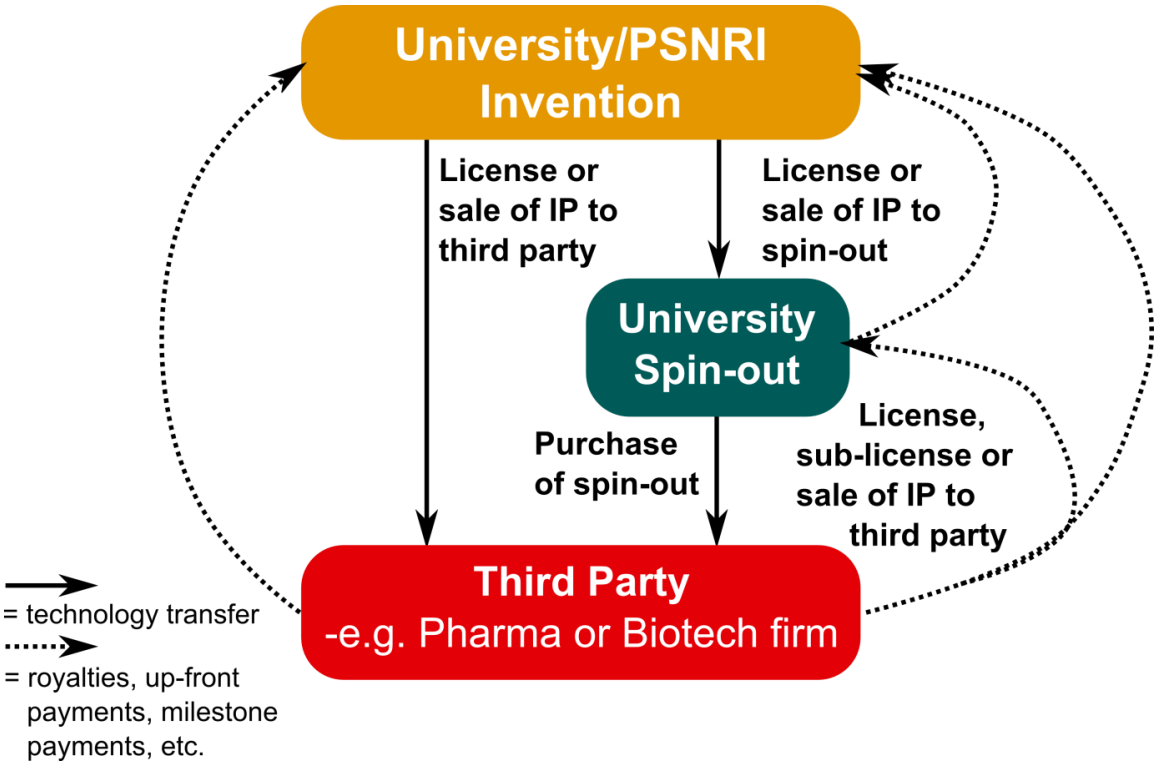


Figure 4.4.2 - scheme illustrating potential commercialisation strategies for university- or PSNRI-originating research. IP is either licensed/sold directly to third parties, or to a spin-out company formed by the inventors' institution. Solid arrows indicate technology transfer, whilst dotted lines indicate returns to the University/PSNRI.

4.4.6 Product development partners

4.4.6.1 The general picture

NTD product development inherently faces the same risks as the development of developed-world diseases. However, as previously mentioned, there is a huge disparity between the collective or individual disease burdens of NTDs and the R&D activity aimed at controlling them (50). Despite this, the funding for global NTD product R&D has increased dramatically over the previous decade, showing gradual increases up until 2009, and steady levels since then. 71.4% of the US $3.045 billion NTD R&D funding in 2011 was in the form of external grants, with a reported $1.5 billion going to academic and private NTD product developers(66).

The largest funders of NTD research are government institutions, presenting as aid agencies or research councils/science and technology (S&T) agencies. They provide around 65-70% of

global funding for NTD research. The other two major sources of funding for NTD research are industry (pharmaceutical and biotechnology) and philanthropy, contributing 17.2% and 18.7%, respectively, of all global funding in 2011(66).

PDPs take a central role in NTD drug development. They received 23% of global grant funding for NTD research in 2011(66). Furthermore, the BVGH 2012 report on the product developer landscape showed that 40% of all drugs and vaccines in current development were being developed using the PDP model(396). Hence, the PDP brokers are another key partner in NTD product development.

4.4.6.2 Funding for trypanosomiasis

4.3% (US $131.7 million) of global funding reported in the G-FINDER report was dedicated to research in diseases caused by kinetoplastids (leishmaniasis, HAT and Chagas disease). Of this amount, 26% was granted to HAT and 18.7% to Chagas disease. A total of roughly US$18 million was spent on R&D for new drugs in 2011(66). The funding landscape for these diseases appears close to the general picture: 73.3% of funding originates in the public sector and 17.3% is donated by philanthropic organisations. The top three funders of research in kinetoplastid diseases are the US National Institutes of Health (NIH), The Bill & Melinda Gates Foundation (Gates Foundation) and The Wellcome Trust. Industry plays a smaller role here, than in other disease areas, contributing only 9.5% of funding, but it is still the third largest source of funding.

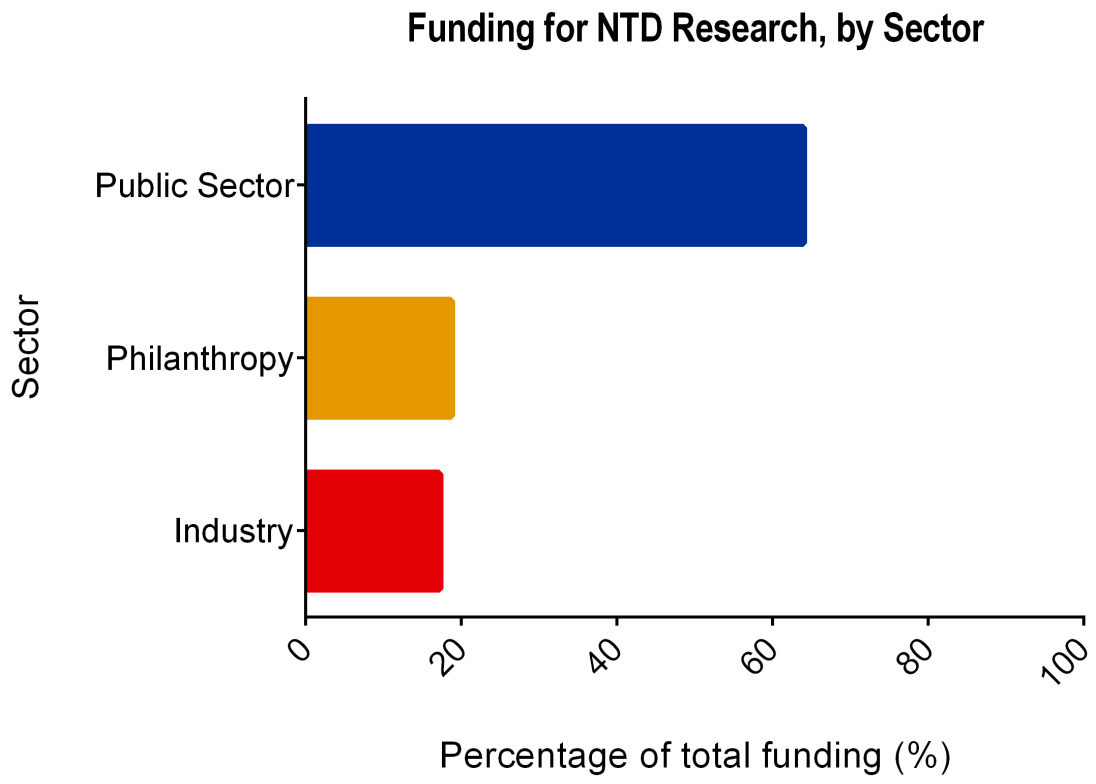


Figure 4.4.3 - global funding for NTD research in 2011, as reported in the G-FINDER report(66). Funding by sector is presented as percentage of total funding.

4.4.6.3 The academic/PSNRI institution as a source of new products for NTDs

In all disease areas, academic institutions and PSNRIs have proven to be important sources for new drugs and vaccines(69). Analysis of some recently-registered NTD products shows academics and PSNRIs at the root of their discoveries(50). A look at the BVGH Global Health Primer also reveals the involvement of such institutions in the development of several NTD products in the pipeline(74). This role in product R&D is particularly important in the NTD setting; the development of innovative treatments is an inherent requirement to treat or prevent several NTDs(50). Evidence to support an innovative role can be found in the literature. In Stevens et al.'s analysis of new drug applications to the US Food and Drug Administration (FDA), they show that academics/PSNRIs were twice as likely as for-profit institutions to meet an unmet medical need when a new drug originated with them. Academic/PSNRI inventors also contributed to 13.6% of new molecular entities (NMEs) included in the analysis, and 9 out of the 10 new drugs that were for new indications(69). Similarly, in Kneller's 2010 analysis of the origins of new drugs, Universities were shown to contribute to the R&D of 30% of drugs given the FDA's priority review (see discussion of PRV, below) and to 31% of drugs considered to be scientifically novel, despite contributing to an estimated 24% of new drugs overall(392).

The prominent role of public funders, coupled with the fact that around three-quarters of all public funding goes directly to researchers(254), means that academic and PSNRIs are collectively the largest recipients of NTD R&D funding. There has also been an increase in intramural funding by PSNRIs and Universities(66). This makes researchers in these positions the most common participants in NTD product development(396). However, groups from Universities and other PSNRIs are often insufficiently resourced to progress new products towards the clinic alone. The costs of preclinical drug development can cost hundreds of thousands of pounds at a minimum, which can be very inhibiting, particularly to a small academic group(394). On the other hand, with the rise of Academic Drug Discovery Centres, particularly in the US, more academic and PSNRI researchers will find the necessary resources available to them(70). Even small groups can collaborate with other groups within their institution or externally in order to carry out different studies using the grants of individual researchers. When looking at the current NTD product pipeline, academic/PSNRIs can be seen to be involved at every stage of the product development lifecycle – from discovery through to clinical phases(74,396).

An academic institution/PSNRI has the option to sell the IP of a particular invention to another product developer to take forward. This is clearly the most convenient option and, in many cases, it is the most appropriate option, relieving researchers of duties that are outside of their area of expertise. There are situations where the development of a product will benefit from the continued input of the original inventors, in which case it would be a good option to license the product to third parties and co-develop the product(397). In addition to these strategies, academics/PSNRIs may spin-out a company to commercialise a product. This has the added value of being able to attract more funds in the form of equity. Being able to provide more funds

directly, means that the inventors take on more of the risk, but are in a more favourable position to reap the commercial rewards of any products that meet the market. Thus, the decision to follow either of the last two strategies is largely dependent on the commercial value of the product being developed. Universities often pursue the spin-out model as a more lucrative revenue source(71,397). However, in the NTD space, most products would have little or no commercial value. This is particularly the case for HAT and it is probably the case for Chagas disease.

The contribution of academic/PSNRIs in overall drug discovery and development, beyond the discovery phase is confirmed by numerous sources(69–71,392). However, in the vast majority of cases, there is some form of technology transfer from Universities/PSRIs to Pharmaceutical or Biotechnology companies, who complete late-stage development, and carry out registration, marketing and distribution. In Kneller's study of the origins of drugs, 8% of drugs originated from Universities that transferred their technology to Pharmaceutical companies, whilst 16% originated from Universities that transferred their technology to Biotechnology companies.

4.4.6.4 Partners for NTD product development

PDP Brokers

As previously described, the successes of several NTD drug and vaccine development programs have been attributed to the PDP model(72). The PDP brokers play a central role in this and manage much of the global funding that is directed to NTD research and development (see above). The characteristics of these brokers differ slightly. There are those that focus on a particular project, whilst there are others that are focused on a particular disease or group of diseases (e.g. Medicines for Malaria Venture [MMV], Global TB Alliance) and manage several products. Brokers such as DNDi and BVGH work across therapeutic areas in the NTD space. As the Brokers have extensive networks and are present at every stage of pharmaceutical development, it would be advantageous to engage a PDP early on in the process. The multi-project and multi-disease Brokers have the added advantage that they have more experience, as they tend to have existed for longer, and thus they are better connected. One disadvantage of PDPs is that they do not generate their own income and rely on funding from external sources. Theoretically, this would make them vulnerable to the same inconsistency in funding as academics/PSRIs. However, this does not seem to be the case, judging by the literature and media reports.

According to the BVGH Global Health Primer, DNDi and the Global Alliance for TB Drug Development, along with BVGH, are PDP brokers that are currently involved in trypanosomiasis drug development.

Government

The biggest proportion of global R&D funding for NTD research comes from government sources. A lot of this funding is given in pursuit of international development goals or for scientific reasons. According to the G-FINDER report, the US NIH (granting a huge US$1.16 billion over five years) is by far the most significant funder of NTD research. Researchers in the UK have access to funds from the European Commission (EC), UK DFID and The Medical Research Council, which are all in the top twelve funders of NTD research. Aid agencies and Research Councils/S&T agencies of the US, Dutch and Australian governments are also present in the list(66). There has been a shift over the years towards more basic research, with 31.2% of funding being dedicated to this area. Therefore, researchers are likely to find funds from these sources for discovery and early-stage development. However, a substantial proportion of funds are still available for later development. For example, in 2011-2012, the MRC released a Highlight Notice, prioritising translational research in NTDs(398). This remains a priority for the MRC's Infection and Immunity Board(399). Funds from research councils/S&T agencies are usually disbursed through competitive grant systems, where researchers must submit project proposals in response to calls for applications.

A large part of the support from government aid agencies focuses on implementation of mass treatment administration programs. This certainly seems to be the main role of USAID(400). The UK DFID is also heavily involved in this area, but also places a lot of emphasis on research to combat NTDs(401). Their contributions are delivered through the WHO TDR initiative and DNDi programs(402)

Philanthropy

Philanthropic donations for NTD research do not seem to be concentrated in any particular disease area. This funding usually plays a contributing role to specific projects and is focused on product development. Around half of all PDP funding comes from philanthropic organisations (principally The Gates Foundation)(66). Thus, it is important to highlight that, although Philanthropists are not as prominent as governments in NTD research funding overall, they are often major sources of funding for the development of individual products. The two largest philanthropic funders are The Gates Foundation, providing an average of US$120.3 million annually in 2007-2011, and The Wellcome Trust, which gave an average of US$10.8 million annually to NTD research over the same period(66). The main areas of focus of The Gates Foundation are Global Health, Global Development, Education (in the US) and Global Policy & Advocacy(403). Efforts to control NTDs receives strong support from the Gates Foundation's Global Health Division and HAT is one of the "High Opportunity Targets" that they focus on(404). Apart from funding PDPs and other global health organisations, The Gates Foundation also has a mechanism for giving grants to researchers. Concepts with specified aims are developed internally, then researchers/organisations are either invited privately or publically to apply for grants in a similar way to how they would apply for research council grants(405). The Wellcome Trust operates in a similar way to research councils. Infectious Diseases are included in the list of 'Research Challenges' in the Trust's 2012-2020 Strategic Plan(406). There are also two specific programs which are applicable to NTD product R&D. Firstly, the Global Health trials scheme provides funding for late-stage clinical trials of 'interventions that will improve health' in low- and middle-income countries(407). Secondly, the Wellcome Trust's Pathfinder Awards provide funding for Academic-Industry partnerships for R&D for NTDs and Orphan Diseases. The aim of this initiative is to advance projects from an early-stage to a stage where there is enough credibility to attract funds from other sources(408). Target validation is not supported by this scheme, so it can be envisaged that this scheme would be attractive to researchers working in the discovery and pre-clinical development stages. Furthermore, similar lines of research could be funded through other, less specific Wellcome Trust Grants.

There is little information on the role of Philanthropy by wealthy individuals in funding NTD research. It is possible that these individuals will give their donations to PDP brokers, and other charities/non-profit organisations, so their contribution is unclear. Given the engagement of such individuals in issues of social importance(409), they should be approached to support R&D for NTD product development.

Large Pharmaceutical Companies

Pharmaceutical firms are one of the primary partners of academic/PSNRI organisations for drug and vaccine development. They are also a key partner in NTD drug development. A search of the G-FINDER database, and a review of the signatories to the London Declaration on NTDs reveals a significant number of large pharmaceutical corporations that are involved in NTD product R&D(43,66). They are responsible for a large part of the industry participation in NTD product development and some companies are engaged in the development of several products at a time(396). In addition to the potentially large amount of funds at their disposal, large pharmaceutical firms also contribute to a large extent with in-kind support, ranging from access to facilities/compounds, offering of expertise, etc.(66). There has been a significant increase in pharmaceutical industry involvement in NTD product development(66), which has been attributed to social and ethical concerns, efforts to improve reputation, and long-term strategic considerations(57).

Much of the increase in funding has been directed towards advancing particular products through late-stage trials. Most of the funding directed towards controlling NTDs is spent internally by large pharmaceutical companies(66), so it is likely that one of the main ways of partnering with them is through some kind of technology transfer of an invention to the company. Much of the partnering done by large pharmaceutical firms is now shifting to earlier stages of development(52,57,66). One of the biggest strengths of this type of partner is that it is one of a few types of organisation that can advance a project from any point in the product life cycle to the market. The 20 or so large, multi-national corporations also have wide networks to engage CROs and other partners to carry out different parts of the development work. Therefore, these firms can be approached for partnership at any stage in development. Of the largest companies, work in the NTD space is dominated by four of them: GlaxoSmithKline, Novartis, Sanofi and Astra Zeneca. Although they have large amounts of money at their disposal, this means that there are really only a small number of individual developers of this type to partner with, compared with the biotechnology cohort.

Biotechnology/Small Pharmaceutical Companies

It has been shown that technology transfer from Universities to biotechnology firms accounts for more new drugs and biologics than that from Universities to pharmaceutical companies (16% vs. 8%) (392). Therefore, biotechnology companies should be considered as one of the primary partners for researchers seeking to develop NTD products. It's often the case that these companies will later partner with or transfer technology to large pharmaceutical companies, but some biotechnology companies will advance their products all the way from discovery to market(392). This highlights the extent of some of their capabilities. Funding from biotech and small-pharmaceutical firms accounted for 10.6% of 2011 funding from industry that was reported in the G-FINDER report, and made up large/majority contributions in some disease areas, such as helminth diseases or trachoma(66).

Compared to the biotechnology sector as a whole, a very small proportion of companies are involved in NTD product development. However, participation by biotechnology companies in NTD product development is estimated by BVGH to be evident across 41% of the NTD product pipeline. Their participation is also present through all phases of the pharmaceutical product development lifecycle, and it exceeds the participation of pharmaceutical companies at all stages of development before phase III. It is possible that biotechnology companies are filling very important roles in different parts of the lifecycle, sometimes where large industry participation is absent(396).

When it is appreciated that a main obstacle in finding funding to develop NTD products is a lack of commercial potential, it is encouraging that biotechnology firms have played a more significant role than large pharmaceutical firms in discovering and developing orphan drugs (drugs developed for rare diseases; see discussion on rare diseases and orphan drugs below), where the commercial value is typically lower(392). This is analogous to the situation that has drawn some small companies to neglected diseases (this was the case for Zentaris in the development of miltefosine, for example) (57). In lower-income countries, where NTDs are present, this may be a further encouragement to engage in NTD product R&D. In 2011, 28.9% of funding reported by biotechnology and small pharmaceutical firms came from countries in lower-income countries. Firms from the emerging markets of India and Brazil are playing a particularly prominent role, and there have been significant increases in research directed towards treating kinetoplastid diseases in Brazilian firms(66).

Biotechnology and other smaller firms are limited by the funds available to them. As their funds originate from investments, as opposed to the revenues of large pharmaceutical companies, they have less flexibility, thus typically contribute less to NTD product development(396). Furthermore, it can be expected that these firms are also less able to offer in-kind contributions than their larger counterparts. PDPs often have to contribute more funding to biotechnology firms, than to pharmaceutical firms, making them a more expensive partner for PDP brokers to work with(72,396).

Contract Research Organisations

Whether conducting neglected disease R&D or not, CROs can be a useful partner for carrying out specific studies that researchers may lack the ability to do in their own facilities. CROs are reasonably active in the field of NTD product R&D and partner extensively with PDPs: the DNDi website lists 28 different CROs from companies from both developed and developing countries as partners(410). An analysis of traceable external PDP expenditure in 2007 shows that 20.9% went to CROs, larger than that which went to large pharmaceutical companies (18%) and biotechnology/small pharmaceutical companies (9.3%) (72).

There are some companies that are combining the role of a traditional pharmaceutical or biotechnology company, with contract services. Founded with money from equity investments,

the South Africa-based iThemba raises additional money through contract medicinal chemistry services and invests the profit into research for treatments of unmet medical needs in Africa. Using this business model, they have successfully generated a pipeline of leads and molecular scaffolds active against malarial parasites, TB and HIV(411). Another company, called Scynexis, has played an important role in developing a late-stage HAT compound(412). This is detailed in a case study, later in this chapter.

Venture Capital

When a researcher's commercialisation strategy is to form a new company, after gaining seed/start-up capital, larger investments are often sought from venture capitalists(397).

Venture Capital (VC) firms provide finance in exchange for some form of equity in the company, but individual deals can take many forms. VC firms usually make an investment with the assumption that there will be a profitable 'exit', at a later date. This will usually involve the company being purchased at a value that is higher than that invested in the company.

In addition to Private VC firms, Corporate Venture Capital (CVC), coming from large corporations, is another source of equity. There has been an increasing trend in venture capital investments to biotechnology/small pharmaceutical firms from large corporations, particularly large pharmaceutical companies. Several large pharmaceutical companies now have venture capital arms. Corporate venturing can be seen as another way that large firms can augment their pipelines and access new technologies. As some private VC firms have moved away from the biotechnology sector, Corporate VC has become increasingly important. Investments are reportedly on the increase, with a shift towards earlier stages of development, but these firms are often looking for companies that are already venture-backed(413).

The lack of financial incentive for making such investments for companies in the NTD space is probably a reason why venture capital has not been looked at as a significant source for funding in this area. However, this does not preclude it from playing a role. There will be a number of cases mentioned in this chapter, where funds from equity investments have eventually played a role in NTD drug development. Additionally, there is already evidence of a venture capital firm engaging in the fight to control neglected diseases. The Legatum group, a Dubai-based private investment firm, is a founder of the END fund, which is a global fund committed to supporting mass drug administration to eradicate NTDs in Africa(414).

Angel Investors

Angels, wealthy individuals that are usually serial entrepreneurs, make investments in a similar manner to venture capitalists, and are another source of funding for new businesses. Like the individual philanthropists, they are individuals who wish to invest in causes that they have some kind of interest in. They often seek to have some influence on the company, too(397). For

reasons already discussed, such an individual is likely to be a philanthropist in the area of NTD product R&D.

4.4.7 Details of partnerships

4.4.7.1 NTD product development in the 21st century

One of the most striking things of Pedrique et al.'s analysis of new products for NTDs (see products in Table 4.4.2), is that there are very few new innovative products – the majority of products are repurposed drugs (denoted NI, for new indication), new fixed dose combinations (FDCs) of existing drugs or new formulations (NF) of existing drugs(52). At first glance, there appears to be a promising number of new vaccines introduced in the period studied. However, these only cover two diseases, and two vaccines originate from the same invention. There are only four NCEs on the whole list(52). As a result, this gives us little information on how a product may be developed from the discovery phase, through to product registration. On the other hand, they do show evidence of successful product development through the more expensive, later stages of development. Therefore, there are several useful lessons to be learned from these successes.

Table 4.4.2 - New NTD products approved or recommended during 2000-11. Table reproduced from the work of Pedrique et al.(52), with an additional column on product development history (from media reports, WHO documents and scientific literature).

Generic Name	Marketed Name(s)	Indication(s)	Formulation	Product Type	Regulatory Body(s)	Country	Date of Approval	MA Holder	Comments on Development History
artemotil		malaria	injectable	NCE	MEB	Netherlands	2000	Artecef BV	Artecef BV was created by ACF Beheer S.A. for the purpose of developing Artemotil with the WHO. Patent held by the WHO Scientific Working Group on the Chemotherapy of Malaria.
atovaquone plus proguanil (paediatric)	Malarone Paediatric	malaria	tablet	NI	FDA	USA	2000	GlaxoSmithKline	
proguanil plus chloroquine	Avloclor	malaria	tablet	FDC	AFSSAPS	France	2001	AstraZeneca	
chloroquine plus primaquine		malaria	tablet	FDC	ANVISA	Brazil	2001	Industria quimica do Estado de Goias S/A - Iquedo, Brazil	
artesunate		malaria	injectable	NF	ANVISA	Brazil	2001	Silvestre Labs Quimica e Farmaceutica Ltda, Brazil	
artemether plus lumefantrine	Coartem	malaria	dispersible tablet	NF	Swissmedic	Switzerland	2008	Novartis Pharma Schweiz AG	Ciba-Geigy, later Novartis entered into a partnership with the Academy of Military Medical Sciences in China and several private Chinese entities to develop Coartem. Originally marketed as Riamet for developed-world travellers, but was produced non-profitably

									later, following an agreement with the WHO. Supported by the Global Fund.
artesunate plus amodiaquine (ASAQ)	Coarsucam	malaria	tablet	FDC	WHO PQ		2008	Sanofi-Aventis Group	Combination developed in parallel by Sanofi and the FACT (Fixed Dose Artesunate Combination Therapy) Consortium - a group headed by DNDi, and involving Farmanguinhos/Fiocruz and various other academic, clinical and CRO partners. They later teamed up to co-develop the product. Compound registration and manufacture was also initiated in Morocco, to build capacity and aid registration in endemic countries.
artesunate plus mefloquine (ASMQ)		malaria	tablet	FDC	ANVISA	Brazil	2008	Farmanguinhos/ Fiocruz, Brazil	Combination devised by WHO TDR staff, with support from USAID and the Wellcome Trust. Developed in a PDP led by the FACT consortium. A later technology transfer agreement allowed production and registration in India.
artemether plus lumefantrine		malaria	oral suspension	NF	CDSCO	India	2008	Information unavailable	
arterolane maleate plus piperaquine phosphate	Synriam	malaria	tablet	NCE	CDSCO	India	2011	Ranbaxy labs	Invented in a PDP, led by MMV, with UK, US and Swiss contributors. MMV licensed the IP to Ranbaxy in 2003.

piperaquine tetraphosphate plus alpha dihydroartemisinin	Eurartesim	malaria	tablet	NCE	EMEA	Europe	2011	Sigma-Tau Industrie Farmaceutiche Riunite SpA	Combination initially developed by Professor Li Guoqiao of Guangzhou University of Traditional Chinese Medicine and developed in a partnership with the Holley group, China. Further developed for Global use in a PDP involving MMV.
artesunate		malaria	suppository	NF	ANVISA	Brazil	2000	Nova Quimica Farmaceutica Ltda, Brazil	Supporting research was funded by WHO TDR, several governments and two PSNRIs.
ethambutol plus isoniazid plus pyrazinamide plus rifampicin		tuberculosis	tablet	FDC	WHO PQ		2003	Lupin Ltd, India	Combination was introduced on advice of WHO Advisory committee.
ethambutol plus isoniazid plus rifampicin		tuberculosis	tablet	FDC	WHO PQ		2008	Macleods Pharmaceuticals Ltd, India	Application submitted by MSF and The International Union against Tuberculosis and Lung Disease.
isoniazid plus rifampicin		tuberculosis	dispersible tablet	NF	WHO PQ		2009	Macleods Pharmaceuticals Ltd, India	
isoniazid plus pyrazinamide plus rifampicin		tuberculosis	dispersible tablet	NF	WHO PQ		2009	Macleods Pharmaceuticals Ltd, India	
moxifloxacin		tuberculosis	tablet	NI	WHO PQ		2010	Cipla Ltd, India	Use is currently off-label, although recommended by WHO; Bayer (after withdrawing its patent in India) and TB alliance are conducting trials to register the product.
levofloxacin		tuberculosis	tablet	NI	WHO PQ		2011	Cipla Ltd, India	
ofloxacin		tuberculosis	tablet	NI	WHO PQ		2011	Cipla Ltd, India	

benznidazole (paediatric)		Chagas disease	dispersible tablet	NF	ANVISA	Brazil	2011	Pernambuco State Pharmaceutical Laboratory (LAFEPE), Brazil	Partnership with DNDi and multiple funders to produce new formulation. DNDi supported product registration.
miltefosine	Impavido	visceral leishmaniasis	capsule	NI	CDSCO	India	2002	German Remedies Ltd, Mumbai	WHO TDR worked closely with Zentaris AG (original inventor) to clinically develop miltefosine for visceral leishmaniasis. WHO TDR set up a development team and several Indian principal investigators were involved. Strong agreements on patient access were put in place. Zentaris licensed it to German Remedies for manufacture and Distribution.
paromomycin		visceral leishmaniasis	injectable	NI	CDSCO	India	2006	Gland Pharma United and One World Health (San Francisco [IOWH]) and the Bill & Melinda Gates Foundation	
miltefosine	Impavido	cutaneous leishmaniasis		NI	CDSCO	India	2008	German Remedies Ltd, Mumbai	Zentaris appear to have sponsored further trials in this indication independently of TDR.
nitazoxanide	Alinia	cryptosporidiosis/ giardiasis	oral suspension	NF	FDA	USA	2002	Romark Laboratories, LC	See below
nitazoxanide	Alinia	cryptosporidiosis/ giardiasis	tablet	NCE	FDA	USA	2004	Romark Laboratories, LC	Originally discovered by Jean-Francois Rossignol in the Pasteur Institute.

									Rossignol co-founded a Romarck to commercialise nitazoxanide.
tosufloxacin tosilate hydrate	Ozex	cholera	granules	NF	PMDA	Japan	2009	Toyama Chemical Co, Ltd	A reformulation of Ozex tablets.
bivalent inactivated vaccine, killed whole cells of *Vibrio cholerae* O1 and O139	Shanchol	cholera	oral suspension	vaccine	CDSCO	India	2009	Shantha Biotech	International Vaccine Institute altered mORCVAX formulation of Vabiotech to meet International standards, with backing from the Gates Foundation and various international development agencies. The technology was then transferred to Shantha for production.
live pentavalent vaccine, for prevention of G1, G2, G3, G4, and G-serotypes containing P1A[8], in Vero cells	RotaTeq	rotavirus	oral suspension	vaccine	EMEA	Europe	2006	Sanofi Pasteur MSD, SNC	Discovered by the Philadelphia Childrens Hospital and the Wistar Institute. Product licensed to Merck and Royalty Interest of Philadelphia Children's Hospital and part royalty interest of Wistar institute was later sold to 3rd parties.
Live, attenuated vaccine, rotavirus RIX4414 strain, for prevention of G1 and non-G1 serotypes (G3, G4, and G9), in Vero cells	Rotarix	rotavirus	oral suspension	vaccine	EMEA	Europe	2006	GlaxoSmithKline Biologicals SA	Licensed from Avant to GSK. This vaccine was first registered for use in Mexico and the Dominican Republic in 2004, and has been approved in more than 35 countries and the European Union.

Japanese encephalitis virus vaccine, freeze-dried, inactivated, Beijing-1 strain, Vero cell-derived	ENCEVAC	Japanese encephalitis	injectable	vaccine	PMDA	Japan	2006	Kaketsuken	
Japanese encephalitis virus vaccine, live chimeric viral vector	Imojev	Japanese encephalitis	injectable	vaccine	TGA	Australia	2008	Sanofi Pasteur Pty Ltd	Developed by Acambis as ChimeriVax-JE. Marketing and distribution license granted to Sanofi. Target markets in South-East Asia, presumably for-profit.
Japanese encephalitis virus vaccine, purified, formalin-inactivated, whole virus vaccine strain SA14-14-2, in Vero cells	Ixiaro	Japanese encephalitis	injectable	vaccine	EMEA	Europe	2009	Intercell AG	Intercell licensed relevant IP from Vaccgen International LLC and Sanofi Pasteur. Distribution by Novartis.
Japanese encephalitis virus vaccine, live, attenuated strain SA14-14-2, in primary hamster kidney cell cultures	mORCVAX	Japanese encephalitis	injectable	vaccine	CDSCO	India	2010	Curevin Pharma Pvt	Curevin are providing Vabiotech's vaccine (see Shanchol, above).
Crotalidae polyvalent immune fab (ovine)		snakebite	injectable	biologic (anti-venom)	FDA	USA	2000	Protherics Inc.	Developed for North American market.
nifurtimox (combination therapy with eflornithine		human African trypanosomiasis	tablet	NI	WHO EML		2009	Drugs for Neglected Diseases Initiative (DNDi)	Combination supported by research conducted by MSF and epicentre.
zinc		diarrhoea	tablet and liquid	NI	WHO EML		2006	Newborn and Child	Research performed by academics, supported by

						Development Department of Child and Adolescent Health, WHO	NGOs.
ribavirin	viral haemorrhagic fevers	intravenous and oral	NI	WHO EML	2007	Biorisk Reduction for Dangerous Pathogens Team (BDP), WHO	

Regarding the products listed in Table 4.4.2, it can be seen that a variety of different organisations filed for registration and hold marketing authorisations (MA) (i.e. are authorised to market and sell the product). The majority is made up of pharmaceutical manufacturers, mainly large pharmaceutical corporations and smaller firms in the industry. There are also some PDP brokers and transnational organisations, serving a similar role to PDP brokers, in the list: DNDi, the Newborn and Child Development Department of Child and Adolescent Health of the WHO, and the Biorisk Reduction for Dangerous Pathogens Team of the WHO.

The aminoglycoside antibiotic, paromomycin was launched in 2006 as a treatment for visceral leishmaniasis (VL) with joint marketing authorisation going to Gland Pharma United, the Institute for One World Health (IOWH) and the Gates Foundation. IOWH has a role as a PDP broker, but also carries out R&D activities. Originally discovered in the 1950s, paromomycin was shown to have efficacy in the treatment of VL in the early 1960s. There was a long history of use and development of paromomycin by trans-national NGOs and public health authorities, MSF and WHO TDR, but the discontinuation of paromomycin production and the cessation of development almost prevented this treatment from having a place in VL therapy. In 1999, IOWH began completion of paromomycin development and received funding from the Gates Foundation to realise this goal. They partnered with Indian pharmaceutical company, Gland Pharma, to ensure production(415,416). This is an example of a successful PDP, which highlights a number of key strategies. IOWH is a non-profit developer, so they were able to develop a treatment where a significant commercial incentive did not exist. They were able to do this, only with philanthropic support. It was also important to engage a producer in a country where VL was endemic.

The case of paromomycin highlights another important pattern seen in Table 4.4.2: 18 of the 33 industrial MA holders are based in developing countries where NTDs are prevalent, India and Brazil. These two countries are also the two largest developing country funders of NTD research(66).

Most of the products in Table 4.4.2 were developed in a partnership, or there was some form of technology transfer between different entities. Another trend seen here is that almost all products that are marketed in lower- or middle-income country markets and/or on a not-for-profit basis received some form of public, philanthropic or NGO support in their development. The WHO or PDPs were involved in supporting research (see artemotil, Synriam®, ASMQ, miltefosine and others) and product registrations (see benznidazole [paediatric]). State-owned pharmaceutical firms were involved in the development of several products in Brazil, and academic/PSNRIs were the home institutions of the original inventors of several of the products (e.g. artemisinin derivatives, such as artemether and artesunate). Even in some cases, where a product was originally targeted towards developed country markets (e.g. Coartem® for travellers), engagement by one of the aforementioned organisations (the WHO in the case of Coartem®) was the key factor in enabling patient access in lower-income countries.

The results of the upsurge in NTD drug development and the emergence of new strategies have not yet been seen in full. The analysis above provides a somewhat misleading picture of the current state of NTD drug development. The BVGH global health primers shows pharmaceutical development projects at every stage in the development life cycle. In March 2012, 190 drugs and 225 drugs were reported as being in 'active development' for 23 different NTDs(417). Thus it is important to analyse some of the partnerships of drugs in development. Fexinidazole, a new treatment for HAT is in late clinical trials and is a great example of the effectiveness of the PDP strategy for developing NTD products. Case studies of other products, which have required the use of innovative business models or strategies, are discussed near the end of this chapter.

Fexinidazole

Fexinidazole is being developed in a PDP driven by DNDi in partnership with Sanofi, the Swiss Tropical Health Institute and HAT platform partners, a consortium of at least 18 national HAT control programs, research institutes and NGOs(74,418). Fexinidazole, a broad spectrum nitroimidazole antibiotic, was in preclinical development to treat bacterial infections in the 1970s, before efficacy against *T. brucei* infection was discovered in the 1980s. It was not developed further for either indication, but it was re-discovered during a large characterisation of 700+ compounds, initiated by DNDi in 2005(419,420). DNDi was responsible for much of the R&D: preclinical, clinical and pharmaceutical. Sanofi carried out industrial development, registration and manufacturing. Members of the HAT platform, particularly the national control programs of the Democratic Republic of the Congo (DRC) and the Central African Republic (CAR), along with Médecins Sans Frontières (MSF) were responsible for clinical trials. DNDi list a number of funders, ranging from the Gates Foundation to government international aid departments(418). Fexinidazole is expected to be available on the market in 2014(421).

4.4.8 Benefits of different options

The presence of different potential partners at different stages of the development lifecycle means that several different approaches can be taken to commercialisation of research in this area. Whether the IP protecting an invention is sold, licensed, or used as a foundation for forming a new company, partnering with other organisations is essential. This is evident when looking at the development histories of most drugs for NTDs(50,57,72) (see Table 4.4.2), and is a key characteristic of pharmaceutical development when the IP originates from academic/PSNRI labs(392).

Chesbrough & Schwartz describe how an effective product co-development partnership should look to leverage the capabilities of the respective partners. In order for such partnerships to be successful, the business objectives of each partner should be aligned(422). An example of business objectives for the commercialisation of the research presented in this thesis are:

- Gain evidence of the efficacy of compounds that are effective in killing trypanosomes

- Optimise leads to increase efficacy
- Prove *in vitro* and *in vivo* safety of new leads
- Prove in-human efficacy of new anti-trypanosomal drug (clinical trials)
- Register new treatment for HAT/Chagas Disease
- Large-scale manufacture and distribution of new drug for HAT/Chagas Disease

Perhaps, a reason for the success of the partnership models used in NTD product development is the alignment of these objectives. Different partners can contribute to projects in a complimentary manner, utilising capabilities that are required for the success of the project, but without perceiving any threats to their core business model. For example, a reason for the collaboration by large pharmaceutical companies in early stages of R&D (i.e. discovery) is that projects are pre-competitive at that stage – the outcomes do not bring the collaborators into competition with each other.

The main questions for academic/PSNRI researchers to address when trying to commercialise research that could lead to a new NTD product, revolve around where to find funding, who to partner with and at what stage(s).

4.4.8.1 The most appropriate option?

Table 4.4.3 - Market segmentation matrix, adapted from G Moore's 'Crossing The Chasm'(423). This tool is usually used to identify an initial market segment for a new product launch. In a similar manner, here it is used to identify the most appropriate type of organisation to partner with. A higher score indicates a more favourable option.

Key Criteria for Partner Selection	Segment 1 **Research Councils and Philanthropists**	Segment 2 **PDP Brokers**	Segment 3 **Large Pharmaceutical Companies**	Segment 4 **Biotechnology Companies**
Is the partner well-funded? Are they readily accessible?	4	3	5	3
Do they have a compelling reason to partner with us?	5	5	3	2
Can we, with the help of the partner, deliver a whole product to the market?	2	5	5	3
Is there no entrenched competition that could prevent us from getting a fair shot?	3	3	4	4
If we enter into partnership, will their references allow us to leverage entry into additional partnerships with others?	4	5	4	3
Total Score (5=high/certainly, 3=medium, 1=low/unsure)	18	21	21	15

Using Moore's market segment rating grid(423), it is possible to compare the relative strengths of potential partners (or 'market segments'). PDP brokers and large pharmaceutical firms appear as the best initial partners. In reality, partnership with one type of third party does not preclude a partnership with another. The analysis does highlight PDP brokers as the most appropriate partner for a number of reasons. They have access to funds, they have extensive networks with which to find partners and they have project management capabilities. Large pharmaceutical companies share many of these capabilities. However, PDP brokers have a business model that is designed specifically for the development of NTD products. A large pharmaceutical company may actually be limited by their business model, which requires profit,

so a third party may need to be engaged to support non-profitable efforts, as has been the case with some NTD products developed in partnership with them.

From an academic/PSNRI researcher's perspective, an important advantage of engaging with one of the two preferred partners is that it relieves them of the significant effort of securing funds for several different development activities and then constructing partnership agreements with numerous third parties. One concern in "handing off" projects, particularly when IP is sold or licensed without further participation from the inventors, is a loss of control over the progression of the product's development. Large pharmaceutical companies are sometimes forced to discontinue the development of a project for financial reasons(72). In this respect, either engaging a PDP or ensuring that partnership agreements involve continued participation would be a safer option.

In NTD product R&D, royalties are generally not a concern for academics/PSNRIs. However, this would be more of a concern if a new company was formed around the IP. For the institution as a whole, forming a new company and taking equity in it can have a significantly higher return than licensing to third parties(424). Given the desire for profitability for traditional start-up funders (i.e. venture capitalists) it would not be advisable to follow this strategy, in the absence of a product that can become profitable.

From a public health perspective, the most favourable outcome is an effective new drug to treat trypanosomiasis that is made available to the target population within the shortest time possible, and at the lowest cost to all parties involved. It is clear that the available evidence strongly points to the PDP model as the most effective approach to achieve this(57,72).

4.4.9 Evidence for Partnerships

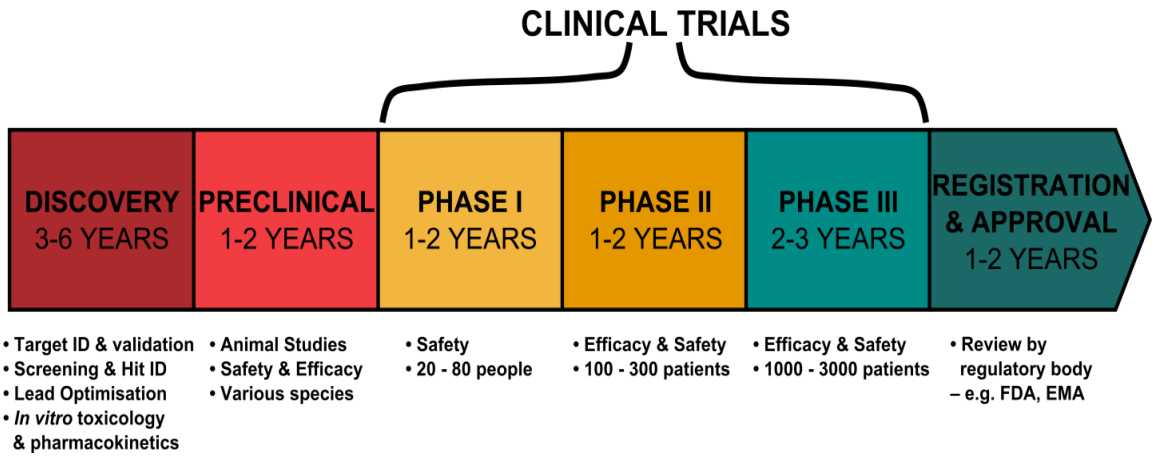


Figure 4.4.4 - Summary of the pharmaceutical development lifecycle, from discovery to registration and approval. Timelines and patient numbers informed by DiMasi et al.(425), S Blank(426) and the FDA(427).

4.4.9.1 Technology transfer in the pharmaceutical industry

Co-development partnerships are not unique to the NTD field, where pharmaceuticals are concerned. When seeking a partner, it is important to understand their motivations for entering

a partnership, as well as anything that may discourage a potential partner from entering into an agreement. An analysis of the evidence to support a particular project is warranted.

Pharmaceutical development can be an incredibly costly and risky process. An analysis of drug development success rates showed that only 19% of drugs entering clinical trials between 1993 and 2004 eventually attained clinical approval. This number was a little higher for co-developed (in-licensed or out-licensed drugs), than for drugs developed solely by one of the 50 pharmaceutical companies analysed in the study (27% vs. 16%). This may indicate that the licensors of new drugs usually prove efficacy at phase I or phase II, before licensing out to a third party. Encouragingly, systemic anti-infective drugs had a higher success rate of 24%, particularly at the transition from phase II to phase III development. This may be due to clearer clinical end points(428). Achieving the dual objectives of forming an effective partnership and eventually gaining clinical approval rely heavily on gaining evidence to support your project.

In an analysis of biologic drug licensing deals from biotechnology to large pharmaceutical companies, Kalamus and Pinkus report that the majority of deals occur between the preclinical phase and phase II. However, phase II appears to be the preferred time(429).

In their analysis, Kalamas and Pinkus show that both parties could generate value by entering into agreements earlier. Under-pricing by large pharmaceutical companies was stated as a cause of delayed deal making(429). With this problem removed in the case of not-for profit endeavours, deals should be made earlier, so that researchers can start to exploit the capabilities of their better-resourced partners sooner.

It has been reported that, after a shift towards deals made after in-human proof-of-concept, early-stage pharmaceutical industry deals are re-emerging: they represented more than 60% of 'big buyer' deals in 2011-2012. Large pharmaceutical companies are also forming more early-stage partnerships directly with academia. These partnerships often involve payments at milestones, when key research objectives have been met. They are sometimes tied to 'options' where, at the completion of all research objectives (i.e. phase II trials), the company can acquire the IP, or even an IP-holding company (where a small company is involved). This kind of arrangement allows the large companies to reduce their exposure to risk, whilst still enabling researchers to advance projects(430).

4.4.9.2 When to partner and with what supporting evidence

The most favourable time to start partnering largely depends on the commercialisation strategy followed by the researchers. In the absence of commercial considerations, it would even be advantageous to seek partners at the discovery phase. If an academic/PSNRI researcher is developing a product that may have some commercial value, then they may wish to attempt technology transfer later in the development lifecycle. In either case, researchers in an academic/PSNRI or in a small company may reach what is often termed "The Valley of Death" or the "Darwinian Sea", which is the point where more funds are required to develop a product,

but there is a lack of willing partners to fund the development of a product, which may still carry a high risk of failure(431,432). As highlighted by Kalamas & Pinkus, having the opportunity to use the capabilities of a well-resourced partner, early in the development process, should also be seen as an advantage(429).

The vast sums of money and significant amount of labour required to optimise products and demonstrate their safety can be out of the reach of most academic/PSNRI researchers(70). It can cost hundreds of thousands of dollars to outsource labour-intensive tasks, and *in vitro* and *in vivo* toxicology studies can cost over US$ 1million(394). Animal studies, which can be very expensive, can play a vital role in demonstrating efficacy(431). In these situations, an alternative to licensing to industry is to partner with a well-resourced academic drug discovery centre to carry out some of this work(70). There are also a number of programs to support translational research originating from academic/PSNRI research. One such program is the NIH's Therapeutics for Rare and Neglected Diseases program (TRND). TRND aims to collaborate with researchers of rare and neglected diseases "with the goal of moving promising therapeutics into human trials"(433).

Aside from gaining access to funding, there are other advantages to partnering with larger partners (e.g. pharmaceutical companies) which are due to the partner's greater experience. A review of strategic alliances between biotech companies and pharmaceutical companies between 1988 and 2000, gives some clues as to what these advantages are. It was shown that in the more complex trials of phase II and phase III, a drug is significantly more likely to proceed to the next phase, if the licensee (e.g. large pharmaceutical company) has more experience. In phase I trials, which are less complex, there is only a small advantage in this respect(434).

An advantage of being affiliated to academic institutions or PSNRI is that the capabilities (e.g. biological and clinical expertise) of the institution can be made use of at several points in the lifecycle, as has been done in the development of other products for NTDs or low-income markets (435). Many academic institutions are now engaged in translational research and there are several established 'drug discovery centres' at institutions across the US(70,71). However, once large scale trials are needed, these institutions become limited, so one should seek to partner with a PDP broker and/or large pharmaceutical company, and an organisation experienced in public health (e.g. WHO, MSF).

There are now several studies that can be used to 'de-risk' drug development projects(436). They are often used in a standardised way in the pharmaceutical industry, but it is often said that, in respect to these studies, expectations held by licensees in large pharmaceutical companies and their licensors, are mismatched. There is little specific information on these expectations and how they affect the partnering decisions of funders of NTD product R&D. Thus, this will be the subject of future research.

The reader is referred to the appendix for more details on alternative strategies for developing drugs for NTDs.

4.4.10 The NTD R&D landscape

A dramatic increase in awareness and advocacy for NTD product R&D has been witnessed over the last decade. This is evident in the increased activity in the area and the evolution in product development and commercialisation strategies. Public-Private partnerships have always been a feature of NTD product R&D, but the emergence of the PDP as the primary model for product development is welcome progress. Unfortunately, reports of funding inconsistencies and research gaps for some diseases persist(52,396). Following the global financial crisis, there has been a decrease in funding from philanthropic sources and some government aid agencies. This is of concern because they are the principal funders of PDPs(66). Despite these concerns, researchers working towards developing products for NTDs are in a better position to do so than a decade ago. In addition to the increased funding and awareness, the experience of others is providing evidence of several new ways to attract funding to the cause. Researchers and other stakeholders of NTD product development should also look to the future, as new incentives, such as tax credits(437,438) and Medical R&D treaties(439), could result from recent proposals.

Limitations of this analysis

This research has aimed at providing insight into the main considerations to be taken account of when developing a treatment for a NTD. Thus, many aspects of pharmaceutical R&D, and how it is carried out for NTDs, have been broadly reviewed. However, in reality, some concerns, such as the specifics of licensing deals, intellectual property and target product profiles all demand deep analyses of their own. These concerns have all been eluded to here, but deeper analyses have focused on the most important and immediate problems faced when advancing this research. Specific evidence on the best stage of development to transfer technology was found to be lacking, so this will be the subject of future research.

4.4.10.1 How to commercialise this research

Opinion on the best course of action

Given the therapeutic area, the most appropriate mechanism for commercialising the results of this research would be through licensing to external partners. In the absence of other, profitable indications, the effort and seed capital required to found a spin-out company would not be worth it. If the drug/series of compounds showed efficacy in diseases with a market, then commercialisation along those lines should be carried out separately through licensing of the

original IP. Profits from those endeavours could be used to support development of the original product.

A large, well-resourced partner, such as a PDP broker or large pharmaceutical company, may be sought for partnership as early as possible to make use of their capabilities and accelerate development. Ownership of IP and research objectives should be agreed upon and aligned as early as possible, to prevent any issues once co-development starts.

PDP brokers seem to engage with researchers once a viable product (e.g. a lead compound with some *in vitro* efficacy) has been found. Therefore, with this research, a partnership with a pharmaceutical or biotechnology firm should be used to bring the research to this point.

The greatest possible number of funding sources, government incentives and other support should be taken advantage of. The options will be different, depending on the commercialisation route followed; different options are available to non-profit organisations than to for-profit organisations. Drawing support from numerous sources is advantageous and may prove to be essential to advance a product all the way to the market.

4.4.11 Development of a drug for a neglected tropical disease

With the emergent interest in combating neglected tropical diseases (NTDs), one can find numerous collaborators for drug discovery and development of NTD drugs. Academia, industry and not-for-profits are all represented amongst the cohort of NTD stakeholders.

The work presented in this thesis is an example of academic drug discovery in the early stages of the drug development lifecycle. In order to complete the necessary experiments, numerous collaborations were formed both within UCL and outside of the institution. First of all, access to relevant resources and enabling technologies was provided by funding from the University's technology transfer company, UCL Business, and by working with colleagues at UCL Chemibank and the UCL Institute for Liver and Digestive Health. As stated above, the trypanosome growth inhibition assays carried out by Professor Kelly at the LSHTM were essential parts of this study. Future work could also be driven by similar collaborations. For example, NMR and mass spectrometry facilities are available for the analysis of inhibitors in the UCL Department of Chemistry and at the UCL Royal Free Campus.

Despite the achievements demonstrated so far, many challenges are faced when carrying out drug discovery in an academic environment. Drug discovery inherently involves the collection of vast amounts of data. Without automation, in the form of plate readers and liquid handling equipment, etc., this data collection can be very lengthy and laborious. Additionally, there is a requirement for certain stages of the drug discovery process to be completed within a timely manner. Therefore, resourcing drug discovery, in terms of material, labour and finances, can be a challenge in academia. Industrial partners tend to have more resources, are better prepared to carry out drug discovery and often make use of automation for the large-scale collection of data. Moreover, industrial partners, particularly large pharmaceutical companies, are more capable of sharing their capabilities in-kind. Overall, drug discovery experience and expertise is concentrated in industry. There are now several companies in the pharmaceutical industry that are active in NTD drug development and they are an excellent choice of partner for drug discovery.

To finance drug discovery, there are numerous grants available from government and not for profit institutions. This can help to advance development to a later stage, where more opportunities for partnership and more funds may be available.

5. Acknowledgements

I would like to state my gratitude for the collaborative efforts of Professor Jon Kelly at the London School of Hygiene and Tropical Medicine, and Professor Dave Selwood and Dr Edith Chan at UCL ChemiBank. Without their contributions, several of the investigations presented herein would not have been possible. The supply of investigative compounds from the UCL School of Pharmacy by Dr Stephen Hilton is also gratefully acknowledged. I am very grateful to Dr Sally Oxenford for the organic synthesis of compound 4011502.

I would like to thank UCL Business for the financial support provided for this research. The advice and oversight given by Dr Richard Angell and Mr Bill Lindsay is also gratefully acknowledged. The work on the characterisation of known TDH inhibitors and of the newly discovered inhibitors was enabled by having access to the FLUOstar Omega plate reader. Therefore, I wish to thank Dr Nathan Davies, at the department of Liver and Digestive Health at UCL, for granting this access and for offering guidance on the use of the equipment. Some of the research included in this thesis was also made possible by the generous supply of investigative compounds from the Developmental Therapeutics Program of the NCI/NIH. The research around the development of therapeutics for NTDs presented in the Introduction (see Section 1.2) and Discussion (see Sections 4.4.4 to 4.4.10; the research is presented in a single chapter in the appendix) was made possible by an Enterprise Scholarship from UCL Advances.

I am extremely thankful for the support and advice given by my primary supervisor, Professor Jon Cooper. He has been a very willing and excellent mentor throughout my studies. Similarly, I owe thanks to Dr Jon Wilden, my secondary supervisor, for his support and tutorage. I would like to thank every member of the Laboratory for Protein Crystallography for their help, particularly Dr Peter Erskine, who contributed significantly during this project. Finally, I am very grateful to my Family and friends who have supported and encouraged me during these challenging four years.

6. References

1. Lichtenberg FR. Pharmaceutical Innovation, Mortality Reduction, and Economic Growth. Natl Bur Econ Res Work Pap Ser [Internet]. 1998;No. 6569. Available from: http://www.nber.org/papers/w6569
2. Grootendorst P, Piérard E, Shim M. Life-expectancy gains from pharmaceutical drugs: a critical appraisal of the literature. Expert Rev Pharmacoecon Outcomes Res. 2009 Aug;9(4):353–64.
3. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov. 2004 Aug 1;3(8):711.
4. Sciences NI of GM. Structure-Based Drug Design Fact Sheet [Internet]. [cited 2012 Feb 28]. Available from: http://www.nigms.nih.gov/Education/structure_drugs.htm
5. Drews J. Drug Discovery: A Historical Perspective. Science. 2000 Mar 17;287(5460):1960–4.
6. Drews J. Strategic trends in the drug industry. Drug Discov Today. 2003 May 1;8(9):411–20.
7. Butcher EC. Can cell systems biology rescue drug discovery? Nat Rev Drug Discov. 2005 Jun;4(6):461–7.
8. Matthias A. Integrating research and development: the emergence of rational drug design in the pharmaceutical industry. Stud Hist Philos Sci Part C Stud Hist Philos Biol Biomed Sci. 2005 Sep;36(3):513–37.
9. Blaney J. A very short history of structure-based design: how did we get here and where do we need to go? J Comput Aided Mol Des. 2011 Dec 11;26(1):13–4.
10. Hopkins AL, Groom CR. The druggable genome. Nat Rev Drug Discov. 2002 Sep 1;1(9):727.
11. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The Genome of the African Trypanosome Trypanosoma brucei. Science. 2005 Jul 15;309(5733):416–22.
12. Hotez PJ, Molyneux DH, Fenwick A, Kumaresan J, Sachs SE, Sachs JD, et al. Control of neglected tropical diseases. N Engl J Med. 2007 Sep 6;357(10):1018–27.
13. Pink R, Hudson A, Mouries M-A, Bendig M. Opportunities and Challenges in Antiparasitic Drug Discovery. Nat Rev Drug Discov. 2005;4(9):727–40.
14. Smith C. Drug target validation: Hitting the target. Nature. 2003 Mar 20;422(6929):341–7.
15. Deschamps JR. The role of crystallography in drug design. AAPS J. 2005 Dec;7(4):E813–9.
16. Carvalho AL, Trincão J, Romão MJ. X-Ray Crystallography in Drug Discovery. In: Roque ACA, editor. Ligand-Macromolecular Interactions in Drug Discovery [Internet]. Totowa, NJ: Humana Press; 2010 [cited 2012 Feb 28]. p. 31–56. Available from: http://www.springerlink.com.libproxy.ucl.ac.uk/content/wt007m512463wv72/#section=751007&page=1&locus=88
17. Parker MW. Protein Structure from X-Ray Diffraction. J Biol Phys. 2003;29(4):341–62.
18. Hunter WN. A structure-based approach to drug discovery; crystallography and implications for the development of antiparasite drugs. Parasitology. 1997;114 Suppl:S17–29.
19. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. Discovering High-Affinity Ligands for Proteins: SAR by NMR. Science. 1996 Nov 29;274(5292):1531–4.
20. Rees DC, Congreve M, Murray CW, Carr R. Fragment-based lead discovery. Nat Rev Drug Discov. 2004 Aug;3(8):660–72.

21. Baker M. Fragment-based lead discovery grows up. Nat Rev Drug Discov. 2013 Jan;12(1):5–7.
22. Murray CW, Verdonk ML, Rees DC. Experiences in fragment-based drug discovery. Trends Pharmacol Sci. 2012 Jan 5;33(5):224–32.
23. Phatak SS, Stephan CC, Cavasotto CN. High-throughput and *in silico* screenings in drug discovery. Expert Opin Drug Discov. 2009 Sep;4(9):947–59.
24. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009 Dec 1;30(16):2785–91.
25. Caporuscio F, Rastelli G, Imbriano C, Del Rio A. Structure-Based Design of Potent Aromatase Inhibitors by High-Throughput Docking. J Med Chem. 2011 Jun 23;54(12):4006–17.
26. Leach AR, Green DVS. Computational chemistry in lead identification, library design and lead optimisation. Mol Simul. 2001;26(1):33–49.
27. Onodera K, Satou K, Hirota H. Evaluations of Molecular Docking Programs for Virtual Screening. J Chem Inf Model. 2007 Jul 1;47(4):1609–18.
28. Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, et al. A Critical Assessment of Docking Programs and Scoring Functions. J Med Chem. 2005;49(20):5912–31.
29. Sousa SF, Fernandes PA, Ramos MJ. Protein–ligand docking: Current status and future challenges. Proteins Struct Funct Bioinforma. 2006 Oct 1;65(1):15–26.
30. Bajorath J. Integration of virtual and high-throughput screening. Nat Rev Drug Discov. 2002 Nov;1(11):882–94.
31. Perez-Pineiro R, Burgos A, Jones DC, Andrew LC, Rodriguez H, Suarez M, et al. Development of a Novel Virtual Screening Cascade Protocol to Identify Potential Trypanothione Reductase Inhibitors. J Med Chem. 2009 Mar 26;52(6):1670–80.
32. Frearson JA, Wyatt PG, Gilbert IH, Fairlamb AH. Target assessment for antiparasitic drug discovery. Trends Parasitol. 2007 Dec;23(12):589–95.
33. Keller TH, Pichota A, Yin Z. A practical view of “druggability.” Curr Opin Chem Biol. 2006 Aug;10(4):357–61.
34. Owens J. Determining druggability. Nat Rev Drug Discov. 2007 Mar;6(3):187–187.
35. Liu T, Altman RB. Identifying Druggable Targets by Protein Microenvironments Matching: Application to Transcription Factors. CPT Pharmacomet Syst Pharmacol. 2014 Jan 22;3(1):e93.
36. Wellcome Trust. Neglected tropical diseases: a new handle on old problems [Internet]. [cited 2012 Feb 28]. Available from: http://www.wellcome.ac.uk/News/2012/Features/WTVM054006.htm
37. Cavalli A, Lizzi F, Bongarzone S, Belluti F, Piazzi L, Bolognesi ML. Complementary medicinal chemistry-driven strategies toward new antitrypanosomal and antileishmanial lead drug candidates. FEMS Immunol Med Microbiol. 2010 Feb 1;58(1):51–60.
38. Hotez PJ, Pecoul B. “Manifesto” for Advancing the Control and Elimination of Neglected Tropical Diseases. PLoS Negl Trop Dis. 2010 May 25;4(5).
39. WHO | TDR at a glance [Internet]. WHO. [cited 2012 Mar 1]. Available from: http://www.who.int/tdr/publications/about-tdr/strategy/tdr-glance/en/index.html
40. WHO | CONTROL OF NEGLECTED TROPICAL DISEASES IS FEASIBLE [Internet]. WHO. [cited 2012 Mar 1]. Available from: http://www.who.int/neglected_diseases/en/
41. Hotez PJ, Fenwick A, Savioli L, Molyneux DH. Rescuing the bottom billion through control of neglected tropical diseases. The Lancet. 2009 May 2;373(9674):1570–5.
42. Ridley DB, Grabowski HG, Moe JL. Developing Drugs For Developing Countries. Health Aff (Millwood). 2006 Mar 1;25(2):313–24.

43. Endorsements of the London Declaration | Uniting to Combat NTDs [Internet]. [cited 2014 Jan 31]. Available from: http://www.unitingtocombatntds.org/endorsements
44. WHO | Diseases covered by NTD Department [Internet]. WHO. [cited 2012 Mar 1]. Available from: http://www.who.int/neglected_diseases/diseases/en/
45. Hotez PJ. Empowering Women and Improving Female Reproductive Health through Control of Neglected Tropical Diseases. PLoS Negl Trop Dis. 2009 Nov 24;3(11).
46. How to Cure 1 Billion People?--Defeat Neglected Tropical Diseases: Scientific American [Internet]. [cited 2012 Feb 28]. Available from: http://www.scientificamerican.com/article.cfm?id=a-plan-to-defeat-neglected-tropical-diseases
47. University of Washington I. GBD Compare [Internet]. IHME. [cited 2014 Sep 9]. Available from: http://viz.healthmetricsandevaluation.org/gbd-compare/
48. WHO | The 17 neglected tropical diseases [Internet]. WHO. [cited 2013 Dec 10]. Available from: http://www.who.int/neglected_diseases/diseases/en/
49. PLOS Neglected Tropical Diseases: A Peer-Reviewed Open-Access Journal [Internet]. [cited 2014 Jan 15]. Available from: http://www.plosntds.org/static/scope
50. Trouiller P, Olliaro P, Torreele E, Orbinski J, Laing R, Ford N. Drug development for neglected diseases: a deficient market and a public-health policy failure. The Lancet. 2002 Jun 22;359(9324):2188–94.
51. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2012 Dec 15;380(9859):2197–223.
52. Pedrique B, Strub-Wourgaft N, Some C, Olliaro P, Trouiller P, Ford N, et al. The drug and vaccine landscape for neglected diseases (2000–11): a systematic assessment. Lancet Glob Health [Internet]. 2013 Oct [cited 2013 Nov 12]; Available from: http://linkinghub.elsevier.com/retrieve/pii/S2214109X13700780
53. WHO | Chagas disease (American trypanosomiasis) [Internet]. [cited 2010 Dec 2]. Available from: http://www.who.int/mediacentre/factsheets/fs340/en/index.html
54. Simarro PP, Jannin J, Cattand P. Eliminating Human African Trypanosomiasis: Where Do We Stand and What Comes Next? PLoS Med. 2008 Feb 26;5(2):e55.
55. Adhoc Committee (of the WHO) on health research relating to future intervention options. Investing in health research and development [Internet]. Geneva: WHO; 1996 [cited 2014 Feb 4]. Available from: http://www.who.int/tdr/publications/tdr-research-publications/investing-in-health/en/
56. Fehr A, Thürmann P, Razum O. Expert Delphi survey on research and development into drugs for neglected diseases. BMC Health Serv Res. 2011 Nov 16;11(1):312.
57. Moran M. A Breakthrough in R&D for Neglected Diseases: New Ways to Get the Drugs We Need. PLoS Med. 2005 Sep 8;2(9):e302.
58. The PLoS Medicine Editors. The Neglected Diseases Section in PLoS Medicine: Moving Beyond Tropical Infections. PLoS Med. 2008 Feb 26;5(2):e59.
59. Berndt ER, Glennerster R, Kremer MR, Lee J, Levine R, Weizsäcker G, et al. Advance market commitments for vaccines against neglected diseases: estimating costs and effectiveness. Health Econ. 2007;16(5):491–511.
60. Stefanakis R, Robertson AS, Ponder EL, Moree M. Analysis of Neglected Tropical Disease Drug and Vaccine Development Pipelines to Predict Issuance of FDA Priority Review Vouchers over the Next Decade. PLoS Negl Trop Dis. 2012 Oct 25;6(10):e1803.
61. Kwabena Tetteh E. Advance Market Commitments for R&D in Diseases That Disproportionately Affect Low-Income Countries. J World Intellect Prop. 2012;15(4):280–303.

62. Leoni PL. Advance Market Commitment: Some Issues and a Bad Remedy [Internet]. Rochester, NY: Social Science Research Network; 2012 Jul [cited 2013 Nov 30]. Report No.: ID 2104110. Available from: http://papers.ssrn.com/abstract=2104110
63. Hughes B. Priority voucher flops. Nat Biotechnol. 2011 Nov;29(11):958–958.
64. Robertson AS, Stefanakis R, Joseph D, Moree M. The Impact of the US Priority Review Voucher on Private-Sector Investment in Global Health Research and Development. PLoS Negl Trop Dis. 2012 Aug 28;6(8):e1750.
65. Pratt B, Loff B. Linking research to global health equity: The contribution of product development partnerships to access to medicines and research capacity building. Am J Public Health. 2013;103(11):1968–78.
66. George Institute. Global Funding of Innovation for Neglected Diseases (G-FINDER) [Internet]. The George Institute for Global Health. 2013 [cited 2013 Nov 30]. Available from: http://www.georgeinstitute.org/projects/global-funding-of-innovation-for-neglected-diseases-g-finder
67. Morel CM. Reaching Maturity — 25 Years of the TDR. Parasitol Today. 2000 Dec;16(12):522–8.
68. WHO | TDR at a glance [Internet]. WHO. [cited 2012 Mar 1]. Available from: http://www.who.int/tdr/publications/about-tdr/strategy/tdr-glance/en/index.html
69. Stevens AJ, Jensen JJ, Wyller K, Kilgore PC, Chatterjee S, Rohrbaugh ML. The Role of Public-Sector Research in the Discovery of Drugs and Vaccines. N Engl J Med. 2011;364(6):535–41.
70. Frye S, Crosby M, Edwards T, Juliano R. US academic drug discovery. Nat Rev Drug Discov. 2011 Jun;10(6):409–10.
71. Coles LD, Cloyd JC. The Role of Academic Institutions in the Development of Drugs for Rare and Neglected Diseases. Clin Pharmacol Ther. 2012 Aug;92(2):193–202.
72. Moran M, Guzman J, Ropars AL, Illmer A. The role of Product Development Partnerships in research and development for neglected diseases. Int Health. 2010 Jun 1;2(2):114–22.
73. Institute G. Global Funding of Innovation for Neglected Diseases (G-FINDER) database open to public [Internet]. The George Institute for Global Health. 2010 [cited 2014 Jan 15]. Available from: http://www.georgeinstitute.org/media-releases/global-funding-of-innovation-for-neglected-diseases-g-finder-database-open-to-public
74. BIO Ventures for Global Health. Global Health Primer [Internet]. [cited 2014 Jan 15]. Available from: http://www.bvgh.org/Biopharmaceutical-Solutions/Global-Health-Primer.aspx
75. WHO | African trypanosomiasis (sleeping sickness) [Internet]. [cited 2010 Dec 2]. Available from: http://www.who.int/mediacentre/factsheets/fs259/en/
76. Brun R, Blum J, Chappuis F, Burri C. Human African trypanosomiasis. The Lancet. 2010 Jan 9;375(9709):148–59.
77. Van den Bossche P, De Deken R, Brandt J, Geerts S, Geysen D, Berkvens D. The transmission of mixed Trypanosoma brucei brucei/T. congolense infections by tsetse (Glossina morsitans morsitans). Vet Parasitol. 2004 Jan 30;119(2-3):147–53.
78. CDC-Centers for Disease Control and Prevention. African Trypanosomiasis - Biology [Internet]. [cited 2014 Sep 8]. Available from: http://www.cdc.gov/parasites/sleepingsickness/biology.html
79. Lee SH. A fatty-acid synthesis mechanism specialized for parasitism. Nat Rev Microbiol. 2007 Apr 1;5(4):287.
80. Rico E, Rojas F, Mony BM, Szoor B, MacGregor P, Matthews KR. Bloodstream form pre-adaptation to the tsetse fly in Trypanosoma brucei. Front Cell Infect Microbiol [Internet]. 2013 Nov 14 [cited 2014 Sep 8];3. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3827541/

81. MacGregor P, Savill NJ, Hall D, Matthews KR. Transmission Stages Dominate Trypanosome Within-Host Dynamics during Chronic Infections. Cell Host Microbe. 2011 Apr;9:310–8.
82. Kennedy PGE. Human African trypanosomiasis of the CNS: current issues and challenges. J Clin Invest. 2004 Feb 15;113(4):496–504.
83. Lejon V, Franco JR, Simarro PP. African trypanosomiasis [Internet]. [cited 2014 Sep 8]. Available from: http://bestpractice.bmj.com/best-practice/monograph/9999/treatment/details.html
84. Barrett MP, Boykin DW, Brun R, Tidwell RR. Human African trypanosomiasis: pharmacological re-engagement with a neglected disease. Br J Pharmacol. 2007 Dec;152(8):1155–71.
85. Pepin J, Milord F, Guern C, Mpia B, Ethier L, Mansinsa D. Trial of prednisolone for prevention of melarsoprol-induced encephalopathy in gambiense sleeping sickness. Lancet. 1989 Jun 3;1(8649):1246–50.
86. Coura JR. Chagas disease: what is known and what is needed - A background article. Mem Inst Oswaldo Cruz. 2007 Oct;102:113–22.
87. Coura JR, Borges-Pereira J. Chagas disease: 100 years after its discovery. A systemic review. Acta Trop. 2010 Aug;115(1-2):5–13.
88. Buscaglia CA, Campo VA, Frasch ACC, Di Noia JM. Trypanosoma cruzi surface mucins: host-dependent coat diversity. Nat Rev Microbiol. 2006 Mar;4(3):229–36.
89. Lescure F-X, Le Loup G, Freilij H, Develoux M, Paris L, Brutus L, et al. Chagas disease: changes in knowledge and management. Lancet Infect Dis. 2010 Aug;10(8):556–70.
90. CDC-Centers for Disease Control and Prevention. Chagas Disease - Resources for Health Professionals - Antiparasitic Treatment [Internet]. [cited 2014 Sep 7]. Available from: http://www.cdc.gov/parasites/chagas/health_professionals/tx.html
91. Yun O, Lima MA, Ellman T, Chambi W, Castillo S, Flevaud L, et al. Feasibility, Drug Safety, and Effectiveness of Etiological Treatment Programs for Chagas Disease in Honduras, Guatemala, and Bolivia: 10-Year Experience of Médecins Sans Frontières. PLoS Negl Trop Dis. 2009 Jul 7;3(7):e488.
92. CDC-Centers for Disease Control and Prevention. Chagas Disease - Epidemiology & Risk Factors [Internet]. [cited 2014 Sep 6]. Available from: http://www.cdc.gov/parasites/chagas/epi.html
93. Hotez PJ, Dumonteil E, Woc-Colburn L, Serpa JA, Bezek S, Edwards MS, et al. Chagas Disease: "The New HIV/AIDS of the Americas." PLoS Negl Trop Dis. 2012 May 29;6(5):e1498.
94. The Lancet. Chagas' disease—an epidemic that can no longer be ignored. The Lancet. 2006 Aug;368(9536):619.
95. Kennedy PGE. The continuing problem of human African trypanosomiasis (sleeping sickness). Ann Neurol. 2008 Aug 1;64(2):116–26.
96. Nwaka S, Hudson A. Innovative lead discovery strategies for tropical diseases. Nat Rev Drug Discov. 2006 Nov;5(11):941–55.
97. La Greca F, Magez S. Vaccination against trypanosomiasis. Hum Vaccin. 2011 Nov 1;7(11):1225–33.
98. Antigenic variation in African trypanosomiasis: a Memorandum. Bull World Health Organ. 1977;55(6):703–13.
99. Rose WC. Feeding experiments with mixtures of highly purified amino acids. J Biol Chem. 1931 Nov 1;94(1):155–65.
100. McCoy RH, Meyer CE, Rose WC. Feeding experiments with mixtures of highly purified amino acids. J Biol Chem. 1935 Dec 1;112(1):283–302.
101. Ravel JM, Eakin RE, Shive W. Glycine, a precursor of 5(4)-amino-4(5)-imidazolecarboxamide. J Biol Chem. 1948 Jan 1;172(1):67–70.

102. MELTZER HL, SPRINSON DB. The synthesis of 4-C14, N15-L-threonine and a study of its metabolism. J Biol Chem. 1952;197(1):461–74.
103. Chargaff E, Sprinson DB. Studies on the mechanism of deamination of serine and threonine in biological systems. J Biol Chem. 1943 Nov 1;151(1):273–80.
104. Bowyer A, Mikolajek H, Stuart JW, Wood SP, Jamil F, Rashid N, et al. Structure and function of the l-threonine dehydrogenase (TkTDH) from the hyperthermophilic archaeon Thermococcus kodakaraensis. J Struct Biol. 2009 Nov;168(2):294–304.
105. Marcus JP, Dekker EE. Threonine formation via the coupled activity of 2-amino-3-ketobutyrate coenzyme A lyase and threonine dehydrogenase. J Bacteriol. 1993 Oct;175(20):6505–11.
106. Edgar AJ. The human L-threonine 3-dehydrogenase gene is an expressed pseudogene. BMC Genet. 2002 Oct 2;3:18.
107. Lam VMS, Chan IPR, Yeung YG. Role of l-Threonine Deaminase and l-Threonine 3-Dehydrogenase in the Utilization of l-Threonine by Pseudomonas aeruginosa. J Gen Microbiol. 1980 Apr 1;117(2):539–42.
108. Bird MI, Nunn PB. Metabolic homoeostasis of L-threonine in the normally-fed rat. Importance of liver threonine dehydrogenase activity. Biochem J. 1983 Sep 15;214(3):687–94.
109. Edgar AJ. Mice have a transcribed L-threonine aldolase/GLY1 gene, but the human GLY1 gene is a non-processed pseudogene. BMC Genomics. 2005 Mar 9;6(1):32.
110. Elliott WH. A new threonine metabolite. Biochim Biophys Acta. 1958 Aug;29(2):446–7.
111. ELLIOTT WH. Amino-acetone; its isolation and role in metabolism. Nature. 1959 Apr 11;183(4667):1051–2.
112. Neuberger A, Tait GH. The enzymic conversion of threonine to aminoacetone. Biochim Biophys Acta. 1960 Jun 17;41:164–5.
113. Neuberger A, Tait GH. Production of aminoacetone by Rhodopseudomonas spheroides. Biochem J. 1962 Aug;84(2):317–28.
114. Gibson KD, Laver WG, Neuberger A. Initial stages in the biosynthesis of porphyrins. 2. The formation of δ-aminolaevulic acid from glycine and succinyl-coenzyme A by particles from chicken erythrocytes. Biochem J. 1958 Sep;70(1):71–81.
115. McGilvray D, Morris JG. Utilization of l-threonine by a species of Arthrobacter. A novel catabolic role for `aminoacetone synthase'. Biochem J. 1969 May;112(5):657–71.
116. Urata G, Granick S. Biosynthesis of α-Aminoketones and the Metabolism of Aminoacetone. J Biol Chem. 1963 Feb 1;238(2):811–20.
117. Dale RA. Catabolism of threonine in mammals by coupling of l-threonine 3-dehydrogenase with 2-amino-3-oxobutyrate-CoA ligase. Biochim Biophys Acta BBA - Gen Subj. 1978 Dec 18;544(3):496–503.
118. Kavanagh KL, Jörnvall H, Persson B, Oppermann U. Medium- and short-chain dehydrogenase/reductase gene and protein families. Cell Mol Life Sci. 2008 Nov 14;65(24):3895–906.
119. Boylan SA, Dekker EE. L-threonine dehydrogenase of Escherichia coli K-12. Biochem Biophys Res Commun. 1978 Nov 14;85(1):190–7.
120. Boylan SA, Dekker EE. L-threonine dehydrogenase. Purification and properties of the homogeneous enzyme from Escherichia coli K-12. J Biol Chem. 1981 Feb 25;256(4):1809–15.
121. Higashi N, Matsuura T, Nakagawa A, Ishikawa K. Crystallization and preliminary X-ray analysis of hyperthermophilic L-threonine dehydrogenase from the archaeon Pyrococcus horikoshii. Acta Crystallograph Sect F Struct Biol Cryst Commun. 2005 Apr 1;61(Pt 4):432–4.
122. Ishikawa K, Higashi N, Nakamura T, Matsuura T, Nakagawa A. The First Crystal Structure of l-Threonine Dehydrogenase. J Mol Biol. 2007 Feb 23;366(3):857–67.

123. Machielsen R, van der Oost J. Production and characterization of a thermostable L-threonine dehydrogenase from the hyperthermophilic archaeon Pyrococcus furiosus. FEBS J. 2006 Jun;273(12):2722–9.
124. Yoneda K, Sakuraba H, Muraoka I, Oikawa T, Ohshima T. Crystal structure of UDP-galactose 4-epimerase-like l-threonine dehydrogenase belonging to the intermediate short-chain dehydrogenase-reductase superfamily. FEBS J. 2010 Dec 1;277(24):5124–32.
125. Yoneda K, Sakuraba H, Araki T, Ohshima T. Crystal Structure of Binary and Ternary Complexes of Archaeal UDP-galactose 4-Epimerase-like l-Threonine Dehydrogenase from Thermoplasma volcanium. J Biol Chem. 2012 Apr 13;287(16):12966–74.
126. Ueatrongchit T, Asano Y. Highly selective L-threonine 3-dehydrogenase from Cupriavidus necator and its use in determination of L-threonine. Anal Biochem. 2011 Mar 1;410(1):44–56.
127. Schmidt A, Sivaraman J, Li Y, Larocque R, Barbosa JARG, Smith C, et al. Three-Dimensional Structure of 2-Amino-3-ketobutyrate CoA Ligase from Escherichia coli Complexed with a PLP−Substrate Intermediate: Inferred Reaction Mechanism. Biochemistry (Mosc). 2001 May 1;40(17):5151–60.
128. Blackmore MA, Turner JM. Threonine Metabolism via Two-carbon Compounds by Pseudomonas oxalaticus. J Gen Microbiol. 1971 Aug 1;67(2):243–6.
129. Bell SC, Turner JM. Bacterial catabolism of threonine. Threonine degradation initiated by L-threonine-NAD+ oxidoreductase. Biochem J. 1976 May 15;156(2):449–58.
130. Potter R, Kapoor V, Newman EB. Role of threonine dehydrogenase in Escherichia coli threonine degradation. J Bacteriol. 1977;132(2):385–91.
131. Gaitonde MK. Conversion of [U-14C]threonine into 14C-labelled amino acids in the brain of thiamin-deficient rats. Biochem J. 1975 Aug;150(2):285–95.
132. Ballevre O., Houlier M.-L., Prugnaud J., Bayle G., Bercovici D., Seve B., et al. Altered partition of threonine metabolism in pigs by protein-free feeding or starvation. J Physiol - Endocrinol Metab. 1991;24–6.
133. Guerranti R, Pagani R, Neri S, Errico S., Leoncini R, Marinello E. Inhibition and regulation of rat liver L-threonine dehydrogenase by different fatty acids and their derivatives. Biochim Biophys Acta BBA - Gen Subj. 2001 Nov 7;1568(1):45–52.
134. Kritzman G, Okon Y, Chet I, Henis Y. Metabolism of L-Threonine and Its Relationship to Sclerotium Formation in Sclerotium Rolfsii. J Gen Microbiol. 1976 Jul 1;95(1):78–86.
135. Wang J, Alexander P, Wu L, Hammer R, Cleaver O, McKnight SL. Dependence of mouse embryonic stem cells on threonine catabolism. Science. 2009 Jul 24;325(5939):435–9.
136. Alexander PB, Wang J, McKnight SL. Targeted killing of a mammalian cell based upon its specialized metabolic state. Proc Natl Acad Sci. 2011 Sep 20;108(38):15828–33.
137. Shyh-Chang N, Locasale JW, Lyssiotis CA, Zheng Y, Teo RY, Ratanasirintrawoot S, et al. Influence of Threonine Metabolism on S-Adenosylmethionine and Histone Methylation. Science. 2013 Jan 11;339(6116):222–6.
138. Cross GA, Klein RA, Linstead DJ. Utilization of amino acids by Trypanosoma brucei in culture: L-threonine as a precursor for acetate. Parasitology. 1975 Oct;71(2):311–26.
139. Linstead DJ, Klein RA, Cross GAM. Threonine Catabolism in Trypanosoma brucei. J Gen Microbiol. 1977 Aug 1;101(2):243–51.
140. Guler JL, Kriegova E, Smith TK, Lukes J, Englund PT. Mitochondrial fatty acid synthesis is required for normal mitochondrial morphology and function in Trypanosoma brucei. Mol Microbiol. 2008 Mar;67(5):1125–42.
141. Stephens JL, Lee SH, Paul KS, Englund PT. Mitochondrial Fatty Acid Synthesis in Trypanosoma brucei. J Biol Chem. 2007 Feb 16;282(7):4427–36.

142. Gilbert RJ, Klein RA, Miller PGG. The role of threonine in the metabolism of acetyl coenzyme A by Trypanosoma brucei brucei. Comp Biochem Physiol Part B Comp Biochem. 1983;74(2):277–81.
143. Millerioux Y, Ebikeme C, Biran M, Morand P, Bouyssou G, Vincent IM, et al. The threonine degradation pathway of the Trypanosoma brucei procyclic form: the main carbon source for lipid biosynthesis is under metabolic control. Mol Microbiol. 2013 Oct;90(1):114–29.
144. Mazet M, Morand P, Biran M, Bouyssou G, Courtois P, Daulouède S, et al. Revisiting the Central Metabolism of the Bloodstream Forms of Trypanosoma brucei: Production of Acetate in the Mitochondrion Is Essential for Parasite Viability. PLoS Negl Trop Dis. 2013 Dec 19;7(12):e2587.
145. Suong Thu Nguyen, Peter Alexander, Margaret A. Phillips. L-Threonine Dehydrogenase is a non-essential gene in the mammalian blood form Trypanosoma brucei. Molecula Biology Laboraty at Woods Hole; 2011 [cited 2014 Mar 26]. Available from: http://hermes.mbl.edu/mpm/mpm-2011/abstracts/abstract.php?id=139
146. Kilgour V. The electrophoretic mobilities and activities of eleven enzymes of bloodstream and culture forms of trypanosoma brucei compared. Mol Biochem Parasitol. 1980;2(1):51–62.
147. Opperdoes FR, Markoš A, Steiger RF. Localization of malate dehydrogenase, adenylate kinase and glycolytic enzymes in glycosomes and the threonine pathway in the mitochondrion of cultured procyclic trypomastigotes of Trypanosoma brucei. Mol Biochem Parasitol. 1981 Dec 31;4(5–6):291–309.
148. Kaukas A, Gashumba JK, Lanham SM, Dukes P. The substitution of procyclic for bloodstream form Trypanosoma brucei gambiense in isoenzyme studies. Trans R Soc Trop Med Hyg. 1990 Mar;84(2):242–5.
149. Wyllie S, Fairlamb AH. Methylglyoxal metabolism in trypanosomes and leishmania. Semin Cell Dev Biol. 2011 May;22(3):271–7.
150. Jacobs RT, Nare B, Phillips MA. State of the Art in African Trypanosome Drug Discovery. Curr Top Med Chem. 2011 May;11(10):1255–74.
151. Schmidt A, Krauth-Siegel R. Enzymes of the Trypanothione Metabolism as Targets for Antitrypanosomal Drug Development. Curr Top Med Chem. 2002 Nov 1;2(11):1239–59.
152. Myler PJ. Searching the Tritryp genomes for drug targets. Adv Exp Med Biol. 2008;625:133–40.
153. Opperdoes FR, Coombs GH. Metabolism of Leishmania: proven and predicted. Trends Parasitol. 2007 Apr;23(4):149–58.
154. Parimi PS, Gruca LL, Kalhan SC. Metabolism of threonine in newborn infants. Am J Physiol - Endocrinol Metab. 2005 Dec 1;289(6):E981–5.
155. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. 2010 Jan 1;38(suppl 1):D457–62.
156. SDS PAGE [Internet]. [cited 2014 Jun 3]. Available from: http://rcx128.deviantart.com/art/SDS-PAGE-416034318
157. Jakob U, Gaestel M, Engel K, Buchner J. Small heat shock proteins are molecular chaperones. J Biol Chem. 1993 Jan 25;268(3):1517–20.
158. Baneyx F, Mujacic M. Recombinant protein folding and misfolding in Escherichia coli. Nat Biotechnol. 2004 Nov;22(11):1399–408.
159. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. The Proteomics Protocols Handbook [Internet]. Humana Press; 2005 [cited 2014 May 30]. p. 571–607. Available from: http://link.springer.com/protocol/10.1385/1-59259-890-0%3A571

160. Sherwood D, Cooper J. Crystals, X-rays and Proteins: Comprehensive Protein Crystallography. Oxford University Press; 2011. 641 p.
161. Rhodes G. Crystallography made crystal clear: a guide for users of macromolecular models. Academic Press; 2006. 354 p.
162. Incardona M-F, Bourenkov GP, Levik K, Pieritz RA, Popov AN, Svensson O. EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. J Synchrotron Radiat. 2009 Nov 1;16(6):872–9.
163. Powell HR, Johnson O, Leslie AGW. Autoindexing diffraction images with iMosflm. Acta Crystallogr D Biol Crystallogr. 2013 Jul 1;69(Pt 7):1195–203.
164. Winter G. xia2: an expert system for macromolecular crystallography data reduction. J Appl Crystallogr. 2010 Feb 1;43(1):186–90.
165. Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AGW. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr D Biol Crystallogr. 2011 Mar 18;67(4):271–81.
166. Evans P. Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr. 2005 Dec 14;62(1):72–82.
167. Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr D Biol Crystallogr. 2011 Apr 1;67(4):282–92.
168. Stein N, Ballard C. Intensity to amplitude conversion using CTRUNCATE [Internet]. Acta Crystallographica Section A: Foundations of Crystallography. 2009 [cited 2014 Jun 2]. Available from: http://scripts.iucr.org/cgi-bin/paper?a45329
169. Matthews BW. Solvent content of protein crystals. J Mol Biol. 1968 Apr 28;33(2):491–7.
170. Matthews BW. X-ray Crystallographic Studies of Proteins. Annu Rev Phys Chem. 1976;27(1):493–493.
171. Kantardjieff KA, Rupp B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. Protein Sci. 2003 Sep 1;12(9):1865–71.
172. Vagin A, Teplyakov A. MOLREP: an Automated Program for Molecular Replacement. J Appl Crystallogr. 1997 Dec 1;30(6):1022–5.
173. Vagin A, Teplyakov A. Molecular replacement with MOLREP. Acta Crystallogr D Biol Crystallogr. 2009 Dec 21;66(1):22–5.
174. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. *Phaser* crystallographic software. J Appl Crystallogr. 2007 Aug 1;40(4):658–74.
175. Patterson AL. A Fourier Series Method for the Determination of the Components of Interatomic Distances in Crystals. Phys Rev. 1934 Sep 1;46(5):372–6.
176. Patterson AL. A Direct Method for the Determination of the Components of Interatomic Distances in Crystals. Z Für Krist - Cryst Mater. 1935;90(1):517–42.
177. Murshudov GN, Vagin AA, Dodson EJ. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. Acta Crystallogr D Biol Crystallogr. 1997 May 1;53(3):240–55.
178. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004 Nov 26;60(12):2126–32.
179. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010 Mar 24;66(4):486–501.
180. Schüttelkopf AW, van Aalten DMF. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr. 2004 Aug;60(Pt 8):1355–63.
181. BLAST: Basic Local Alignment Search Tool [Internet]. [cited 2012 Apr 12]. Available from: http://blast.ncbi.nlm.nih.gov/Blast.cgi

182. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualisation system for exploratory research and analysis. J Comput Chem. 2004 Oct;25(13):1605–12.
183. Krissinel E, Henrick K. Inference of Macromolecular Assemblies from Crystalline State. J Mol Biol. 2007 Sep 21;372(3):774–97.
184. Pepys MB, Herbert J, Hutchinson WL, Tennent GA, Lachmann HJ, Gallimore JR, et al. Targeted pharmacological depletion of serum amyloid P component for treatment of human amyloidosis. Nature. 2002 May 16;417(6886):254–9.
185. Dunn G, Montgomery MG, Mohammed F, Coker A, Cooper JB, Robertson T, et al. The Structure of the C-C Bond Hydrolase MhpC Provides Insights into its Catalytic Mechanism. J Mol Biol. 2005 Feb 11;346(1):253–65.
186. Andrews P. The gel-filtration behaviour of proteins related to their molecular weights over a wide range. Biochem J. 1965 Sep;96(3):595–606.
187. Sreerama N, Woody RW. Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set. Anal Biochem. 2000 Dec 15;287(2):252–60.
188. Davies GE, Stark GR. Use of Dimethyl Suberimidate, a Cross-Linking Reagent, in Studying the Subunit Structure of Oligomeric Proteins. Proc Natl Acad Sci. 1970 Jul 1;66(3):651–6.
189. Cook PF, Cleland WW. Enzyme Kinetics and Mechanism. London ; New York: Garland Science; 2007. 416 p.
190. Fuwa K, Valle BL. The Physical Basis of Analytical Atomic Absorption Spectrometry. The Pertinence of the Beer-Lambert Law. Anal Chem. 1963;35(8):942–6.
191. McGilvray D, Morris JG. Utilization of L-threonine by a species of Arthrobacter. A novel catabolic role for “aminoacetone synthase”. Biochem J. 1969;112(5):657–71.
192. Hill A. The possible effects of the aggregation of the molecules of haemoglobin on its oxygen dissociation. J Physiol. 1910;
193. Forsén S, Linse S. Cooperativity: over the Hill. Trends Biochem Sci. 1995 Dec;20(12):495–7.
194. Copeland RA. Enzymes: a practical introduction to structure, mechanism, and data analysis. 2nd ed. John Wiley and Sons; 2000. 417 p.
195. Michaelis L, Menten ML. Die kinetik der invertinwirkung. Biochem Z. 1913;49(333-369):352.
196. Johnson KA, Goody RS. The Original Michaelis Constant: Translation of the 1913 Michaelis-Menten Paper. Biochemistry (Mosc). 2011 Oct 4;50(39):8264–9.
197. Kazuoka T, Takigawa S, Arakawa N, Hizukuri Y, Muraoka I, Oikawa T, et al. Novel psychrophilic and thermolabile L-threonine dehydrogenase from psychrophilic Cytophaga sp. strain KUC-1. J Bacteriol. 2003 Aug;185(15):4483–9.
198. Green ML, Elliott WH. The enzymic formation of aminoacetone from threonine and its further metabolism. Biochem J. 1964 Sep;92(3):537–49.
199. Komatsubara S, Murata K, Kisumi M, Chibata I. Threonine degradation by Serratia marcescens. J Bacteriol. 1978 Aug 1;135(2):318–23.
200. Klein RA, Angus JM, Amadife AE, Smith L. Stereospecificity of the threonine dehydrogenase from bloodstream Trypanosoma brucei. Comp Biochem Physiol Part B Comp Biochem. 1980;66(1):143–6.
201. Tressel T, Thompson R, Zieske LR, Menendez MI, Davis L. Interaction between L-threonine dehydrogenase and aminoacetone synthetase and mechanism of aminoacetone production. J Biol Chem. 1986 Dec 15;261(35):16428–37.
202. Ray M, Ray S. L-Threonine dehydrogenase from goat liver. Feedback inhibition by methylglyoxal. J Biol Chem. 1985 May 25;260(10):5913–8.
203. Sittampalam GS, Gal-Edd N, Arkin M, Auld D, Austin C, Bejcek B, et al., editors. Assay Guidance Manual [Internet]. Bethesda (MD): Eli Lilly & Company and the National

Center for Advancing Translational Sciences; 2004 [cited 2014 Jun 2]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK53196/
204. Lamarck JBPA de M de. Zoological philosophy : an exposition with regard to the natural history of animals / by J.B. Lamarck ; translated, with an introduction by Hugh Elliot. London : Macmillan; 1914.
205. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem. 1998 Nov;19(14):1639–62.
206. Using AutoDock 4 with AutoDockTools — AutoDock [Internet]. [cited 2012 Mar 6]. Available from: http://autodock.scripps.edu/faqs-help/tutorial/using-autodock-4-with-autodocktools
207. Which values of the genetic algorithm parameters do you normally use? — AutoDock [Internet]. [cited 2011 Aug 2]. Available from: http://autodock.scripps.edu/faqs-help/faq/which-values-of-the-genetic-algorithm-parameters-do-you-normally-use
208. Hetényi C, van der Spoel D. Efficient docking of peptides to proteins without prior knowledge of the binding site. Protein Sci. 2009 Apr;11(7):1729–37.
209. DTP - Diversity Set Information [Internet]. [cited 2012 Feb 27]. Available from: http://dtp.cancer.gov/branches/dscb/diversity_explanation.html
210. Irwin JJ, Shoichet BK. ZINC − A Free Database of Commercially Available Compounds for Virtual Screening. J Chem Inf Model. 2005 Jan 1;45(1):177–82.
211. Lipkin MJ, Stevens AP, Livingstone DJ, Harris CJ. How Large Does a Compound Screening Collection Need To Be? Comb Chem High Throughput Screen. 2008 Jul;11:482–93.
212. Newman DJ, Cragg GM, Snader KM. Natural Products as Sources of New Drugs over the Period 1981−2002. J Nat Prod. 2011 Nov 3;66(7):1022–37.
213. Newman DJ, Cragg GM. Natural Products as Sources of New Drugs over the Last 25 Years⊥. J Nat Prod. 2007;70(3):461–77.
214. Li JW-H, Vederas JC. Drug Discovery and Natural Products: End of an Era or an Endless Frontier? Science. 2009 Jul 10;325(5937):161–5.
215. Rosén J, Gottfries J, Muresan S, Backlund A, Oprea TI. Novel Chemical Space Exploration via Natural Products. J Med Chem. 2009;52(7):1953–62.
216. Bon RS, Waldmann H. Bioactivity-Guided Navigation of Chemical Space. Acc Chem Res. 2010;43(8):1103–14.
217. Koch MA, Schuffenhauer A, Scheck M, Wetzel S, Casaulta M, Odermatt A, et al. Charting biologically relevant chemical space: A structural classification of natural products (SCONP). Proc Natl Acad Sci U S A. 2005 Nov 29;102(48):17272–7.
218. Camp D, Davis RA, Campitelli M, Ebdon J, Quinn RJ. Drug-like Properties: Guiding Principles for the Design of Natural Product Libraries. J Nat Prod [Internet]. 2011; Available from: http://dx.doi.org/10.1021/np200687v
219. Ogungbe IV, Setzer WN. Comparative Molecular Docking of Antitrypanosomal Natural Products into Multiple Trypanosoma brucei Drug Targets. Molecules. 2009 Apr 14;14:1513–36.
220. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci. 1988;28(1):31–6.
221. Weininger D, Weininger A, Weininger JL. SMILES. 2. Algorithm for generation of unique SMILES notation. J Chem Inf Comput Sci. 1989;29(2):97–101.
222. Weininger D. SMILES. 3. DEPICT. Graphical depiction of chemical structures. J Chem Inf Comput Sci. 1990;30(3):237–43.
223. Daylight Theory: SMILES [Internet]. [cited 2012 Mar 4]. Available from: http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html

224. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH. PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res. 2009 Jun 4;37(Web Server):W623–33.
225. Bolton EE, Wang Y, Thiessen PA, Bryant SH. Chapter 12 PubChem: Integrated Platform of Small Molecules and Biological Activities. Annual Reports in Computational Chemistry [Internet]. Elsevier; 2008 [cited 2012 Mar 4]. p. 217–41. Available from: http://www.sciencedirect.com/science/article/pii/S1574140008000121
226. Molinspiration Cheminformatics [Internet]. [cited 2012 Mar 4]. Available from: http://www.molinspiration.com/
227. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J Cheminformatics. 2011 Oct 7;3(1):33.
228. The Open Babel Package, version 2.3.0 [Internet]. [cited 2011 Aug 7]. Available from: http://openbabel.org/
229. Vainio MJ, Johnson MS. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. J Chem Inf Model. 2007 Nov 1;47(6):2462–74.
230. Cosconati S, Forli S, Perryman AL, Harris R, Goodsell DS, Olson AJ. Virtual screening with AutoDock: theory and practice. Expert Opin Drug Discov. 2010 Jun;5(6):597–607.
231. Steverding D. Parasites & Vectors | Full text | The development of drugs for treatment of sleeping sickness: a historical review. Parasit Vectors [Internet]. 2011 [cited 2011 Sep 27];3(15). Available from: http://www.parasitesandvectors.com/content/3/1/15
232. Koch MA, Wittenberg L-O, Basu S, Jeyaraj DA, Gourzoulidou E, Reinecke K, et al. Compound library development guided by protein structure similarity clustering and natural product structure. Proc Natl Acad Sci U S A. 2004 Nov 30;101(48):16721–6.
233. PDBe < Fold < EMBL-EBI [Internet]. [cited 2013 Jul 4]. Available from: http://www.ebi.ac.uk/msd-srv/ssm/ssmstart.html
234. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr. 2004 Nov 26;60(12):2256–68.
235. Lo Conte L, Ailey B, Hubbard TJP, Brenner SE, Murzin AG, Chothia C. SCOP: a Structural Classification of Proteins database. Nucleic Acids Res. 2000 Jan 1;28(1):257–9.
236. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000 Jan 1;28(1):235–42.
237. RCSB Protein Data Bank - RCSB PDB [Internet]. [cited 2014 Jun 1]. Available from: http://www.rcsb.org/pdb/home/home.do
238. Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, et al. BRENDA, the enzyme information system in 2011. Nucleic Acids Res. 2011 Jan;39(Database issue):D670–6.
239. Google [Internet]. [cited 2014 Jun 1]. Available from: https://www.google.co.uk
240. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res. 2007 Jan 3;35(Database):D198–201.
241. Chang MW, Ayeni C, Breuer S, Torbett BE. Virtual Screening for HIV Protease Inhibitors: A Comparison of AutoDock 4 and Vina. PLoS ONE. 2010 Aug 4;5(8).
242. Maybridge.com [Internet]. [cited 2014 Jun 2]. Available from: http://www.maybridge.com/portal/alias__Rainbow/lang__en/tabID__230/DesktopDefault.aspx
243. Morley AD, Pugliese A, Birchall K, Bower J, Brennan P, Brown N, et al. Fragment-based hit identification: thinking in 3D. Drug Discov Today. 2013 Dec;18(23–24):1221–7.
244. Asinex Libraries [Internet]. [cited 2014 Jun 2]. Available from: http://www.asinex.com/Libraries.html

245. Feng BY, Shelat A, Doman TN, Guy RK, Shoichet BK. High-throughput assays for promiscuous inhibitors. Nat Chem Biol. 2005 Aug;1(3):146–8.
246. Feng BY, Shoichet BK. A Detergent-Based Assay for the Detection of Promiscuous Inhibitors. Nat Protoc. 2006;1(2):550–3.
247. Willett P, Barnard JM, Downs GM. Chemical Similarity Searching. J Chem Inf Comput Sci. 1998 Nov 1;38(6):983–96.
248. Willett P. Similarity-based virtual screening using 2D fingerprints. Drug Discov Today. 2006 Dec;11(23–24):1046–53.
249. Nettles JH, Jenkins JL, Bender A, Deng Z, Davies JW, Glick M. Bridging Chemical and Biological Space: "Target Fishing" Using 2D and 3D Molecular Descriptors. J Med Chem. 2006;49(23):6802–10.
250. Csizmadia F. JChem: Java Applets and Modules Supporting Chemical Database Handling from Web Browsers. J Chem Inf Comput Sci. 2000;40(2):323–4.
251. Chemical Hashed Fingerprints « ChemAxon – cheminformatics platforms and desktop applications [Internet]. [cited 2012 Mar 6]. Available from: http://www.chemaxon.com/jchem/doc/user/fingerprint.html#process
252. Eckert H, Bajorath J. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. Drug Discov Today. 2007 Mar;12(5-6):225–33.
253. ScreenMD « ChemAxon – cheminformatics platforms and desktop applications [Internet]. [cited 2012 Mar 7]. Available from: http://www.chemaxon.com/jchem/doc/user/ScreenMD.html
254. Owen JR, Nabney IT, Medina-Franco JL, López-Vallejo F. Visualisation of Molecular Fingerprints. J Chem Inf Model. 2011 Jul 25;51(7):1552–63.
255. Tipping ME, Lowe D. Shadow targets: A novel algorithm for topographic projections by radial basis functions. Neurocomputing. 1998 Apr 21;19(1–3):211–22.
256. Bolanos-Garcia VM, Davies OR. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. Biochim Biophys Acta. 2006 Sep;1760(9):1304–13.
257. Parsy CB, Chapman CJ, Barnes AC, Robertson JF, Murray A. Two-step method to isolate target recombinant protein from co-purified bacterial contaminant SlyD after immobilised metal affinity chromatography. J Chromatogr B. 2007 Jun 15;853(1–2):314–9.
258. Gouy M, Guindon S, Gascuel O. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. Mol Biol Evol. 2010 Feb 1;27(2):221–4.
259. Yuan JH, Austic RE. Characterization of hepatic L-threonine dehydrogenase of chicken. Comp Biochem Physiol B Biochem Mol Biol. 2001 Aug;130(1):65–73.
260. Kao YC, Davis L. Purification and Structural Characterization of Porcine L-Threonine Dehydrogenase. Protein Expr Purif. 1994 Oct;5(5):423–31.
261. Joernvall H, Persson B, Krook M, Atrian S, Gonzalez-Duarte R, Jeffery J, et al. Short-chain dehydrogenases/reductases (SDR). Biochemistry (Mosc). 1995 May 1;34(18):6003–13.
262. Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. Biochim Biophys Acta BBA - Proteins Proteomics. 2005 Aug 10;1751(2):119–39.
263. Stivala A, Wybrow M, Wirth A, Whisstock JC, Stuckey PJ. Automatic generation of protein structure cartoons with Pro-origami. Bioinformatics. 2011 Dec 1;27(23):3315–6.
264. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983 Dec;22(12):2577–637.
265. Joosten RP, te Beek TAH, Krieger E, Hekkelman ML, Hooft RWW, Schneider R, et al. A series of PDB related databases for everyday needs. Nucleic Acids Res. 2011 Jan 1;39(Database):D411–9.

266. Rao ST, Rossmann MG. Comparison of super-secondary structures in proteins. J Mol Biol. 1973 May 15;76(2):241–56.
267. Rossman MG, Liljas A. Recognition of structural domains in globular proteins. J Mol Biol. 1974 May 5;85(1):177–81.
268. Winn MD. An overview of the CCP4 project in protein crystallography: an example of a collaborative project. J Synchrotron Radiat. 2002 Dec 24;10(1):23–5.
269. Mukherjee JJ, Dekker EE. 2-Amino-3-ketobutyrate CoA ligase of Escherichia coli: stoichiometry of pyridoxal phosphate binding and location of the pyridoxyllysine peptide in the primary structure of the enzyme. Biochim Biophys Acta BBA - Protein Struct Mol Enzymol. 1990 Jan 19;1037(1):24–9.
270. Stierand K, Rarey M. From Modeling to Medicinal Chemistry: Automatic Generation of Two-Dimensional Complex Diagrams. ChemMedChem. 2007 Jun 11;2(6):853–60.
271. Stierand K, Rarey M. Drawing the PDB: Protein−Ligand Complexes in Two Dimensions. ACS Med Chem Lett. 2010 Dec 9;1(9):540–5.
272. Stierand K, Rarey M. PoseView -- molecular interaction patterns at a glance. J Cheminformatics. 2010 May 4;2(Suppl 1):P50.
273. Selwood T, Tang L, Lawrence SH, Anokhina Y, Jaffe EK. Kinetics and Thermodynamics of the Interchange of the Morpheein Forms of Human Porphobilinogen Synthase†. Biochemistry (Mosc). 2008 Mar 1;47(10):3245–57.
274. Strelow J, Dewe W, Iversen PW, Brooks HB, Radding JA, McGee J, et al. Mechanism of Action assays for Enzymes [Internet]. 2012 [cited 2013 Mar 26]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK92001/
275. Cortes A, Cascante M, Cardenas ML, Cornish-Bowden A. Relationships between inhibition constants, inhibitor concentrations for 50% inhibition and types of inhibition: new ways of analysing data. Biochem J. 2001 Jul 1;357(Pt 1):263–8.
276. Morrison JF, Walsh CT. The Behavior and Significance of Slow-Binding Enzyme Inhibitors. In: Meister A, editor. Advances in Enzymology and Related Areas of Molecular Biology [Internet]. John Wiley & Sons, Inc.; 2006 [cited 2013 Apr 15]. p. 201–301. Available from: http://onlinelibrary.wiley.com/doi/10.1002/9780470123072.ch5/summary
277. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 1997 Jan 15;23(1-3):3–25.
278. Congreve M, Carr R, Murray C, Jhoti H. A “Rule of Three” for fragment-based lead discovery? Drug Discov Today. 2003 Oct 1;8(19):876–7.
279. Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. J Med Chem. 2002 Jun 1;45(12):2615–23.
280. eulerAPE: Drawing Area-Proportional Euler and Venn Diagrams Using Ellipses [Internet]. eulerAPE. [cited 2014 Feb 21]. Available from: http://www.eulerdiagrams.org/eulerAPE/index.html
281. Wei X, Gao L, Zhang X, Qian H, Rowan K, Mark D, et al. Introducing Bayesian Thinking to High-Throughput Screening for False-Negative Rate Estimation. J Biomol Screen. 2013 May 29;1087057113491495.
282. Johnson AR, Chen Y-W, Dekker EE. Investigation of a Catalytic Zinc Binding Site in Escherichia coli -Threonine Dehydrogenase by Site-Directed Mutagenesis of Cysteine-38. Arch Biochem Biophys. 1998 Oct 15;358(2):211–21.
283. Persson B, Kallberg Y, Bray JE, Bruford E, Dellaporta SL, Favia AD, et al. The SDR (short-chain dehydrogenase/reductase and related enzymes) nomenclature initiative. Chem Biol Interact. 2009 Mar 16;178(1–3):94–8.

284. Nakano S, Okazaki S, Tokiwa H, Asano Y. Binding of NAD+ and L-Threonine induces Stepwise Structural and Flexibility changes in Cupriavidus Necator L-Threonine Dehydrogenase. J Biol Chem. 2014 Feb 20;jbc.M113.540773.
285. Wagner M, Andreesen JR. Purification and characterization of threonine dehydrogenase from Clostridium sticklandii. Arch Microbiol. 1995 Apr 1;163(4):286–90.
286. Benner SA. The stereoselectivity of alcohol dehydrogenases: A sterochemical imperative? Experientia. 1982 May 1;38(5):633–7.
287. Winberg JO, Brendskag MK, Sylte I, Lindstad RI, McKinley-McKee JS. The catalytic triad in Drosophila alcohol dehydrogenase: pH, temperature and molecular modelling studies. J Mol Biol. 1999 Nov 26;294(2):601–16.
288. Gani OABSM, Adekoya OA, Giurato L, Spyrakis F, Cozzini P, Guccione S, et al. Theoretical Calculations of the Catalytic Triad in Short-Chain Alcohol Dehydrogenases/Reductases. Biophys J. 2008 Feb 15;94(4):1412–27.
289. Koumanov A, Benach J, Atrian S, Gonzàlez-Duarte R, Karshikoff A, Ladenstein R. The catalytic mechanism of Drosophila alcohol dehydrogenase: Evidence for a proton relay modulated by the coupled ionization of the active site Lysine/Tyrosine pair and a NAD+ ribose OH switch. Proteins Struct Funct Bioinforma. 2003;51(2):289–98.
290. Wuxiuer Y, Morgunova E, Cols N, Popov A, Karshikoff A, Sylte I, et al. An intact eight-membered water chain in drosophilid alcohol dehydrogenases is essential for optimal enzyme activity. FEBS J. 2012 Aug 1;279(16):2940–56.
291. Aoyama Y, Motokawa Y. L-Threonine dehydrogenase of chicken liver. Purification, characterization, and physiological significance. J Biol Chem. 1981 Dec 10;256(23):12367–73.
292. Sullivan SM, Holyoak T. Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. Proc Natl Acad Sci. 2008 Sep 16;105(37):13829–34.
293. Nagel ZD, Klinman JP. A 21st century revisionist's view at a turning point in enzymology. Nat Chem Biol. 2009;5(8):543–50.
294. Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci. 2010 Oct;35(10):539–46.
295. Hammes GG, Chang Y-C, Oas TG. Conformational selection or induced fit: A flux description of reaction mechanism. Proc Natl Acad Sci. 2009 Aug 18;106(33):13737–41.
296. Aly A, Edrada-Ebel R, Wray V, Proksch P. Antitrypanosomal metabolites from an endophytic Penicillium sp. isolated from Limonium tubiflorum. Planta Med [Internet]. 2010 Aug [cited 2011 Nov 3];76(12). Available from: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0030-1264700
297. Kurganov BI, Kagan ZS, Dorozhko AI, Yakovlev VA. Kinetic manifestations of allosteric interactions in models of regulatory enzymes with "indirect" co-operativity. J Theor Biol. 1974 Sep;47(1):1–41.
298. Endrenyi L, Kwong FHF, Fajszi C. Evaluation of Hill Slopes and Hill Coefficients when the Saturation Binding or Velocity is not Known. Eur J Biochem. 1975 Feb 1;51(2):317–28.
299. Krystyna B. The role of malic enzyme in the carbohydrate metabolism of Trichinella spiralis spiralis and trichinella spiralis pseudospiralis. Int J Parasitol. 1986 Oct;16(5):435–40.
300. Kurganov BI, Dorozhko AI, Kagan ZS, Yakovlev VA. The theoretical analysis of kinetic behaviour of "hysteretic" allosteric enzymes. II. The dissociating and associating enzymic systems in which the rate of installation of equilibrium between the oligomeric forms is small in comparison with that of enzymatic reaction. J Theor Biol. 1976 Aug 7;60(2):271–86.

301. Kurganov BI. The theoretical analysis of kinetic behaviour of “Hysteretic” allosteric enzymes: IV. Kinetics of dissociation—association processes of allosteric enzymes. J Theor Biol. 1977 Oct 21;68(4):521–43.
302. Weiss JN. The Hill equation revisited: uses and misuses. FASEB J Off Publ Fed Am Soc Exp Biol. 1997 Sep;11(11):835–41.
303. Koshland DE, Némethy G, Filmer D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. Biochemistry (Mosc). 1966;5(1):365–85.
304. MONOD J, WYMAN J, CHANGEUX JP. ON THE NATURE OF ALLOSTERIC TRANSITIONS: A PLAUSIBLE MODEL. J Mol Biol. 1965 May;12:88–118.
305. Selwood T, Jaffe EK. Dynamic dissociating homo-oligomers and the control of protein function. Arch Biochem Biophys. 2012 Mar 15;519(2):131–43.
306. Jaffe EK, Lawrence SH. The morpheein model of allostery: Evaluating proteins as potential morpheeins. Methods Mol Biol Clifton Nj. 2012;796:217–31.
307. Lawrence SH, Jaffe EK. Expanding the concepts in protein structure-function relationships and enzyme kinetics: Teaching using morpheeins. Biochem Mol Biol Educ. 2008 Jul 1;36(4):274–83.
308. Brahma A, Bhattacharyya D. UDP-galactose 4-epimerase from Kluyveromyces fragilis: existence of subunit independent functional site. FEBS Lett. 2004 Nov 5;577(1-2):27–34.
309. Brahma A, Banerjee N, Bhattacharyya D. UDP-galactose 4-epimerase from Kluyveromyces fragilis – catalytic sites of the homodimeric enzyme are functional and regulated. FEBS J. 2009;276(22):6725–40.
310. Walsh R. Alternative Perspectives of Enzyme Kinetic Modeling. In: Ekinci D, editor. Medicinal Chemistry and Drug Design [Internet]. InTech; 2012 [cited 2014 Aug 30]. Available from: http://www.intechopen.com/books/medicinal-chemistry-and-drug-design/alternative-perspectives-of-enzyme-kinetic-modeling
311. Boylan SA, Dekker EE. L-threonine dehydrogenase. Purification and properties of the homogeneous enzyme from Escherichia coli K-12. J Biol Chem. 1981 Feb 25;256(4):1809–15.
312. Boylan SA, Dekker EE. Growth, enzyme levels, and some metabolic properties of an Escherichia coli mutant grown on L-threonine as the sole carbon source. J Bacteriol. 1983 Oct;156(1):273–80.
313. Olsen JV, Ong S-E, Mann M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol Cell Proteomics MCP. 2004 Jun;3(6):608–14.
314. Bechara EJH, Dutra F, Cardoso VES, Sartori A, Olympio KPK, Penatti CAA, et al. The dual face of endogenous α-aminoketones: Pro-oxidizing metabolic weapons. Comp Biochem Physiol Part C Toxicol Pharmacol. 2007 Aug;146(1–2):88–110.
315. Copeland RA, Pompliano DL, Meek TD. Drug–target residence time and its implications for lead optimization. Nat Rev Drug Discov. 2006 Sep;5(9):730–9.
316. Dahl G, Akerud T. Pharmacokinetics and the drug–target residence time concept. Drug Discov Today. 2013 Aug;18(15–16):697–707.
317. Green ML. A time-dependent effect of K+ on l-threonine dehydrogenase. Biochim Biophys Acta BBA - Spec Sect Enzymol Subj. 1963;67:682–5.
318. Green ML. The activation of l-threonine dehydrogenase by potassium ions. Biochem J. 1964 Sep;92(3):550–5.
319. Zhao H. Effect of ions and other compatible solutes on enzyme activity, and its implication for biocatalysis using ionic liquids. J Mol Catal B Enzym. 2005 Dec 1;37(1–6):16–25.
320. Hofmeister F. Zur Lehre von der Wirkung der Salze - Zweite Mittheilung. Arch Für Exp Pathol Pharmakol. 1888;24(4-5):247–60.

321. Kunz W, Henle J, Ninham BW. “Zur Lehre von der Wirkung der Salze” (about the science of the effect of salts): Franz Hofmeister’s historical papers. Curr Opin Colloid Interface Sci. 2004 Aug;9(1–2):19–37.
322. Broering JM, Bommarius AS. Evaluation of Hofmeister Effects on the Kinetic Stability of Proteins. J Phys Chem B. 2005 Nov 1;109(43):20612–9.
323. Cera ED. A Structural Perspective on Enzymes Activated by Monovalent Cations. J Biol Chem. 2006 Jan 20;281(3):1305–8.
324. Kiser PD, Lorimer GH, Palczewski K. Use of thallium to identify monovalent cation binding sites in GroEL. Acta Crystallograph Sect F Struct Biol Cryst Commun. 2009 Oct 1;65(Pt 10):967–71.
325. Villeret V, Huang S, Fromm HJ, Lipscomb WN. Crystallographic evidence for the action of potassium, thallium, and lithium ions on fructose-1,6-bisphosphatase. Proc Natl Acad Sci. 1995 Sep 12;92(19):8916–20.
326. Epperly BR, Dekker EE. L-Threonine Dehydrogenase from Escherichia Coli. Identification of an Active Site Cysteine Residue and Metal Ion Studies. J Biol Chem. 1991 Apr 5;266(10):6086–92.
327. Hassell AM, An G, Bledsoe RK, Bynum JM, Carter HL, Deng S-JJ, et al. Crystallization of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr. 2007 Jan 1;63(Pt 1):72–9.
328. Murphy RM. Static and dynamic light scattering of biological macromolecules: what can we learn? Curr Opin Biotechnol. 1997 Feb;8(1):25–30.
329. Hanlon AD, Larkin MI, Reddick RM. Free-Solution, Label-Free Protein-Protein Interactions Characterized by Dynamic Light Scattering. Biophys J. 2010 Jan 20;98(2):297–304.
330. Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. Transient Protein-Protein Interactions: Structural, Functional, and Network Properties. Structure. 2010 Oct 13;18(10):1233–43.
331. Pierce MM, Raman CS, Nall BT. Isothermal Titration Calorimetry of Protein–Protein Interactions. Methods. 1999 Oct;19(2):213–21.
332. Meneses E, Mittermaier A. Electrostatic interactions in the binding pathway of a transient protein complex studied by NMR and isothermal titration calorimetry. J Biol Chem. 2014 Aug 13;jbc.M114.553354.
333. Cole JL, Lary JW, Moody T, Laue TM. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. Methods Cell Biol. 2008;84:143–79.
334. Lebowitz J, Lewis MS, Schuck P. Modern analytical ultracentrifugation in protein science: A tutorial review. Protein Sci Publ Protein Soc. 2002 Sep;11(9):2067–79.
335. Yee AA, Savchenko A, Ignachenko A, Lukin J, Xu X, Skarina T, et al. NMR and X-ray Crystallography, Complementary Tools in Structural Proteomics of Small Proteins. J Am Chem Soc. 2005;127(47):16512–7.
336. Barrientos LG, Louis JM, Botos I, Mori T, Han Z, O’Keefe BR, et al. The Domain-Swapped Dimer of Cyanovirin-N Is in a Metastable Folded State. Structure. 2002 Jan 5;10(5):673–86.
337. Crowley PB, Ubbink M. Close Encounters of the Transient Kind: Protein Interactions in the Photosynthetic Redox Chain Investigated by NMR Spectroscopy. Acc Chem Res. 2003;36(10):723–30.
338. Tochtrop GP, Richter K, Tang C, Toner JJ, Covey DF, Cistola DP. Energetics by NMR: Site-specific binding in a positively cooperative system. Proc Natl Acad Sci. 2002 Feb 19;99(4):1847–52.
339. Wei S, Kashiwagi M, Kota S, Xie Z, Nagase H, Brew K. Reactive Site Mutations in Tissue Inhibitor of Metalloproteinase-3 Disrupt Inhibition of Matrix Metalloproteinases but Not Tumor Necrosis Factor-α-converting Enzyme. J Biol Chem. 2005 Sep 23;280(38):32877–82.

340. Cole JC, Murray CW, Nissink JWM, Taylor RD, Taylor R. Comparing protein–ligand docking programs is difficult. Proteins Struct Funct Bioinforma. 2005 Aug 15;60(3):325–32.
341. Chen H, Lyne PD, Giordanetto F, Lovell T, Li J. On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors. J Chem Inf Model. 2006 Jan 1;46(1):401–15.
342. Bottegoni G, Rocchia W, Rueda M, Abagyan R, Cavalli A. Systematic Exploitation of Multiple Receptor Conformations for Virtual Ligand Screening. PLoS ONE. 2011 May 17;6(5):e18845.
343. Huang S-Y, Zou X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. Proteins. 2007 Feb 1;66(2):399–421.
344. Craig IR, Essex JW, Spiegel K. Ensemble docking into multiple crystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments. J Chem Inf Model. 2010 Apr 26;50(4):511–24.
345. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol. 2008 Apr;18(2):178–84.
346. Durrant JD, Urbaniak MD, Ferguson MAJ, McCammon JA. Computer-Aided Identification of Trypanosoma brucei Uridine Diphosphate Galactose 4'-Epimerase Inhibitors: Toward the Development of Novel Therapies for African Sleeping Sickness. J Med Chem. 2010 Jul 8;53(13):5025–32.
347. De Graaf C, Pospisil P, Pos W, Folkers G, Vermeulen NPE. Binding Mode Prediction of Cytochrome P450 and Thymidine Kinase Protein−Ligand Complexes by Consideration of Water and Rescoring in Automated Docking. J Med Chem. 2005 Apr 1;48(7):2308–18.
348. Huggins DJ, Venkitaraman AR, Spring DR. Rational Methods for the Selection of Diverse Screening Compounds. ACS Chem Biol. 2011 Mar 18;6(3):208–17.
349. Xie X-Q, Chen J-Z. Data Mining a Small Molecule Drug Screening Representative Subset from NIH PubChem. J Chem Inf Model. 2008;48(3):465–75.
350. Van Hoorn WP, Bell AS. Searching Chemical Space with the Bayesian Idea Generator. J Chem Inf Model. 2009 Oct 26;49(10):2211–20.
351. Brooks WH, McCloskey DE, Daniel KG, Ealick SE, Secrist JA, Waud WR, et al. In Silico Chemical Library Screening and Experimental Validation of a Novel 9-Aminoacridine Based Lead-Inhibitor of Human S-Adenosylmethionine Decarboxylase. J Chem Inf Model. 2007 Sep 1;47(5):1897–905.
352. Freitas RF, Prokopczyk IM, Zottis A, Oliva G, Andricopulo AD, Trevisan MTS, et al. Discovery of novel Trypanosoma cruzi glyceraldehyde-3-phosphate dehydrogenase inhibitors. Bioorg Med Chem. 2009 Mar 15;17(6):2476–82.
353. Fischbach MA, Clardy J. One pathway, many products. Nat Chem Biol. 2007 Jul;3(7):353–5.
354. Evans BE, Rittle KE, Bock MG, DiPardo RM, Freidinger RM, Whitter WL, et al. Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. J Med Chem. 1988;31(12):2235–46.
355. Nicolaou KC, Pfefferkorn JA, Roecker AJ, Cao G-Q, Barluenga S, Mitchell HJ. Natural Product-like Combinatorial Libraries Based on Privileged Structures. 1. General Principles and Solid-Phase Synthesis of Benzopyrans. J Am Chem Soc. 2000;122(41):9939–53.
356. Koch MA, Waldmann H. Protein structure similarity clustering and natural product structure as guiding principles in drug discovery. Drug Discov Today. 2005 Apr 1;10(7):471–83.
357. Zhang M-Q, Wilkinson B. Drug discovery beyond the "rule-of-five." Curr Opin Biotechnol. 2007 Dec;18(6):478–88.

358. Shultz MD. Improving the Plausibility of Success with Inefficient Metrics. ACS Med Chem Lett. 2014 Jan 9;5(1):2–5.
359. Copeland RA. Mechanistic considerations in high-throughput screening. Anal Biochem. 2003 Sep 1;320(1):1–12.
360. Merschjohann K, Sporer F, Steverding D, Wink M. In Vitro Effect of Alkaloids on Bloodstream Forms of Trypanosoma brucei and T. congolense. Planta Med. 2001 Oct;67(7):623–7.
361. Hoet S, Opperdoes F, Brun R, Quetin-Leclercq J. Natural products active against African trypanosomes: a step towards new drugs. Nat Prod Rep. 2004 May 5;21(3):353–64.
362. Trabelsi N, Petit P, Manigand C, Langlois d'Estaintot B, Granier T, Chaudière J, et al. Structural evidence for the inhibition of grape dihydroflavonol 4-reductase by flavonols. Acta Crystallogr D Biol Crystallogr. 2008 Aug;D64(Pt 8):883–91.
363. Muanda FN, Dicko A, Soulimani R. Assessment of polyphenolic compounds, in vitro antioxidant and anti-inflammation properties of Securidaca longepedunculata root barks. C R Biol. 2010;333(9):663–9.
364. Aderbauer B, Clausen P-H, Kershaw O, Melzig MF. In vitro and in vivo trypanocidal effect of lipophilic extracts of medicinal plants from Mali and Burkina Faso. J Ethnopharmacol. 2008 Sep 26;119(2):225–31.
365. Tasdemir D, Kaiser M, Brun R, Yardley V, Schmidt TJ, Tosun F, et al. Antitrypanosomal and Antileishmanial Activities of Flavonoids and Their Analogues: In Vitro, In Vivo, Structure-Activity Relationship, and Quantitative Structure-Activity Relationship Studies. Antimicrob Agents Chemother. 2006 Apr 1;50(4):1352–64.
366. Arioka S, Sakagami M, Uematsu R, Yamaguchi H, Togame H, Takemoto H, et al. Potent inhibitor scaffold against Trypanosoma cruzi trans-sialidase. Bioorg Med Chem. 2010 Feb 15;18(4):1633–40.
367. Ammar Z, Plazolles N, Baltz T, Coustou V. Identification of Trans-Sialidases as a Common Mediator of Endothelial Cell Activation by African Trypanosomes. PLoS Pathog. 2013 Oct 10;9(10):e1003710.
368. Coley AF, Dodson HC, Morris MT, Morris JC. Glycolysis in the African Trypanosome: Targeting Enzymes and Their Subcellular Compartments for Therapeutic Development. Mol Biol Int. 2011 Apr 11;2011:e123702.
369. Dodson HC, Lyda TA, Chambers JW, Morris MT, Christensen KA, Morris JC. Quercetin, a fluorescent bioflavanoid, inhibits Trypanosoma brucei hexokinase 1. Exp Parasitol. 2011 Feb;127(2):423–8.
370. Merschjohann K, Steverding D. In vitro growth inhibition of bloodstream forms of Trypanosoma brucei and Trypanosoma congolense by iron chelators. Kinetoplastid Biol Dis. 2006 Aug 16;5(1):3.
371. Petit P, Granier T, d' Estaintot BL, Manigand C, Bathany K, Schmitter J-M, et al. Crystal Structure of Grape Dihydroflavonol 4-Reductase, a Key Enzyme in Flavonoid Biosynthesis. J Mol Biol. 2007 May 18;368(5):1345–57.
372. Roper JR, Güther MLS, Milne KG, Ferguson MAJ. Galactose metabolism is essential for the African sleeping sickness parasite Trypanosoma brucei. Proc Natl Acad Sci. 2002 Apr 30;99(9):5884–9.
373. Roper JR, Güther MLS, MacRae JI, Prescott AR, Hallyburton I, Acosta-Serrano A, et al. The Suppression of Galactose Metabolism in Procylic Form Trypanosoma brucei Causes Cessation of Cell Growth and Alters Procyclin Glycoprotein Structure and Copy Number. J Biol Chem. 2005 May 20;280(20):19728–36.
374. Urbaniak MD, Turnock DC, Ferguson MAJ. Galactose Starvation in a Bloodstream Form Trypanosoma brucei UDP-Glucose 4'-Epimerase Conditional Null Mutant. Eukaryot Cell. 2006 Nov 1;5(11):1906–13.

375. MacRae JI, Obado SO, Turnock DC, Roper JR, Kierans M, Kelly JM, et al. The suppression of galactose metabolism in Trypanosoma cruzi epimastigotes causes changes in cell surface molecular architecture and cell morphology. Mol Biochem Parasitol. 2006 May;147(1):126–36.
376. Chen G., Wang J. Threonine metabolism and embryonic stem cell self-renewal. Curr Opin Clin Nutr Metab Care. 2014;80–5.
377. Williamson J, Scott-Finnigan TJ. Trypanocidal Activity of Antitumor Antibiotics and Other Metabolic Inhibitors. Antimicrob Agents Chemother. 1978 May;13(5):735–44.
378. Rodrigues RR, Lane JE, Carter CE, Bogitsh BJ, Singh PK, Zimmerman LJ, et al. Chelating agent inhibition of Trypanosoma cruzi epimastigotes In vitro. J Inorg Biochem. 1995 Dec;60(4):277–88.
379. Lane JE, Ribeiro-Rodrigues R, Suarez CC, Bogitsh BJ, Jones MM, Singh PK, et al. In vitro trypanocidal activity of tetraethylthiuram disulfide and sodium diethylamine-N-carbodithioate on Trypanosoma cruzi. Am J Trop Med Hyg. 1996 Sep;55(3):263–6.
380. Mantilla BAS. Caracterização funcional e papel fisiológico da D1-pirrolina-5-carboxilato desidrogenase de Trypanosoma cruzi: uma enzima do metabolismo de prolina. [Internet] [text]. Universidade de São Paulo; 2013 [cited 2014 Sep 3]. Available from: http://www.teses.usp.br/teses/disponiveis/42/42135/tde-20032014-081257/
381. Mantilla BAS. Characterization of the Disulfiram-effect in Trypanosoma Cruzi. 2013 Aug 1 [cited 2014 Sep 3]; Available from: http://www.bv.fapesp.br/en/bolsas/144654/characterization-of-the-disulfiram-effect-in-trypanosoma-cruzi/
382. Anighoro A, Bajorath J, Rastelli G. Polypharmacology: Challenges and Opportunities in Drug Discovery. J Med Chem [Internet]. 2014 Jun 19 [cited 2014 Sep 3]; Available from: http://dx.doi.org/10.1021/jm5006463
383. Tallarida RJ. Drug Synergism: Its Detection and Applications. J Pharmacol Exp Ther. 2001 Sep 1;298(3):865–72.
384. Chou T-C. Drug Combination Studies and Their Synergy Quantification Using the Chou-Talalay Method. Cancer Res. 2010 Jan 15;70(2):440–6.
385. Morphy R, Rankovic Z. Designed Multiple Ligands. An Emerging Drug Discovery Paradigm. J Med Chem. 2005 Oct 1;48(21):6523–43.
386. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol. 2008 Nov;4(11):682–90.
387. Gattrell W, Johnstone C, Patel S, Smith CS, Scheel A, Schindler M. Designed multiple ligands in metabolic disease research: from concept to platform. Drug Discov Today. 2013 Aug;18(15-16):692–6.
388. Hopkins AL, Mason JS, Overington JP. Can we rationally design promiscuous drugs? Curr Opin Struct Biol. 2006 Feb;16(1):127–36.
389. Apsel B, Blair JA, Gonzalez B, Nazif TM, Feldman ME, Aizenstein B, et al. Targeted polypharmacology: discovery of dual inhibitors of tyrosine and phosphoinositide kinases. Nat Chem Biol. 2008 Nov;4(11):691–9.
390. Vainio MJ, Kogej T, Raubacher F, Sadowski J. Scaffold Hopping by Fragment Replacement. J Chem Inf Model. 2013 Jul 22;53(7):1825–35.
391. Pankiewicz KW, Chen L, Petrelli R, Felczak K, Gao G, Bonnac L, et al. Nicotinamide Adenine Dinucleotide Based Therapeutics. Curr Med Chem. 2008 Mar 1;15(7):650–70.
392. Kneller R. The importance of new companies for drug discovery: origins of a decade of new drugs. Nat Rev Drug Discov. 2010 Nov;9(11):867–82.
393. Powers JB. Commercializing Academic Research: Resource Effects on Performance of University Technology Transfer. J High Educ. 2003;74(1):26–50.
394. Freeman MW, Dervan AP. The Path From Bench to Bedside: Considerations Before Starting the Journey. J Investig Med June 2011. 2011;59(5):746–51.

395. Peng X. University spin-offs: Opportunity or challenge? Nat Mater. 2006 Dec;5(12):923–5.
396. BIO Ventures for Global Health. Developing New Drugs & Vaccines for Neglected Diseases of the Poor: The Product Developer Landscape [Internet]. 2012 Mar. Available from: http://www.bvgh.org/LinkClick.aspx?fileticket=h6a0cJK9drg%3d&tabid=91
397. Adams DJ, Sparrow JC. Enterprise for life scientists: developing innovation and entrepeneurship in the biosciences. Bloxham: Scion; 2008.
398. MRC. Neglected Tropical Diseases (NTDs) highlight notice July 2011 [Internet]. [cited 2014 Jan 30]. Available from: http://www.mrc.ac.uk/Fundingopportunities/Highlightnotices/NTDs/MRC008039
399. MRC. IIB programme [Internet]. [cited 2014 Jan 30]. Available from: http://www.mrc.ac.uk/Ourresearch/Boardpanelsgroups/IIB/Programme/index.htm
400. USAID's Neglected Tropical Disease Program: About the NTD Program [Internet]. [cited 2014 Jan 30]. Available from: http://www.neglecteddiseases.gov/about/index.html#support
401. Improving the health of poor people in developing countries - Policy - GOV.UK [Internet]. [cited 2014 Jan 30]. Available from: https://www.gov.uk/government/policies/improving-the-health-of-poor-people-in-developing-countries
402. Preventing and treating neglected tropical diseases - Improving the health of poor people in developing countries - Policies - GOV.UK [Internet]. [cited 2014 Jan 30]. Available from: https://www.gov.uk/government/policies/improving-the-health-of-poor-people-in-developing-countries/supporting-pages/preventing-and-treating-neglected-tropical-diseases
403. What We Do - Bill & Melinda Gates Foundation [Internet]. [cited 2014 Jan 30]. Available from: http://www.gatesfoundation.org/what-we-do
404. Neglected Infectious Diseases - Bill & Melinda Gates Foundation [Internet]. [cited 2014 Jan 30]. Available from: http://www.gatesfoundation.org/What-We-Do/Global-Health/Neglected-Infectious-Diseases
405. How We Make Grants - Bill & Melinda Gates Foundation [Internet]. [cited 2014 Jan 30]. Available from: http://www.gatesfoundation.org/How-We-Work/General-Information/How-We-Make-Grants
406. The Strategic Plan 2010-20 | Wellcome Trust [Internet]. [cited 2014 Jan 30]. Available from: http://www.wellcome.ac.uk/About-us/Strategy/index.htm
407. Global health trials scheme | Wellcome Trust [Internet]. [cited 2014 Jan 30]. Available from: http://www.wellcome.ac.uk/Funding/Biomedical-science/Funding-schemes/Strategic-awards-and-initiatives/WTX059944.htm
408. Pathfinder Awards | Wellcome Trust [Internet]. [cited 2014 Jan 30]. Available from: http://www.wellcome.ac.uk/Funding/Technology-transfer/Awards/Pathfinder-Awards/index.htm?utm_source=Adestra&utm_medium=email&utm_content=Pathfinder+Awards&utm_campaign=12_10_11+TT+news+issue+12+JF&utm_term=a.c.caplan%40adm.leeds.ac.uk
409. Schervish PG. Major donors, major motives: The people and purposes behind major gifts. New Dir Philanthr Fundrais. 2005;2005(47):59–87.
410. DNDi. Partners [Internet]. [cited 2014 Jan 30]. Available from: http://www.dndi.org/partnership/partners.html#CROs
411. iThemba [Internet]. [cited 2014 Jan 30]. Available from: http://www.ithembapharma.com/index.php
412. Drugs for Neglected Diseases Initiative. Oxaborole SCYX-7158 [Internet]. [cited 2014 Feb 4]. Available from: http://www.dndi.org/diseases-projects/portfolio/oxaborole-scyx-7158.html

413. Von Krogh G, Battistini B, Pachidou F, Baschera P. The changing face of corporate venturing in biotechnology. Nat Biotechnol. 2012 Oct;30(10):911–5.
414. The END Fund Launches to Combat Neglected Tropical Diseases in Africa - The Legatum Group [Internet]. [cited 2014 Jan 30]. Available from: http://www.legatum.com/article/The-END-Fund-Launches-to-Combat-Neglected-Tropical-Diseases-in-Africa
415. Kala-azar [Internet]. OneWorld Health. [cited 2013 Nov 28]. Available from: http://www.oneworldhealth.org/kala-azar
416. Davidson RN, Boer M den, Ritmeijer K. Paromomycin. Trans R Soc Trop Med Hyg. 2009 Jul 1;103(7):653–60.
417. BIO Ventures for Global Health. The Global Health Primer 2012 Snapshot [Internet]. The Global Health Primer 2012 Snapshot. [cited 2014 Jan 31]. Available from: http://www.bvgh.org/LinkClick.aspx?fileticket=PzndmHX3BSU%3d&tabid=91
418. Fexinidazole [Internet]. [cited 2013 Nov 28]. Available from: http://www.dndi.org/diseases-projects/portfolio/fexinidazole.html
419. Torreele E, Bourdin Trunz B, Tweats D, Kaiser M, Brun R, Mazué G, et al. Fexinidazole – A New Oral Nitroimidazole Drug Candidate Entering Clinical Development for the Treatment of Sleeping Sickness. PLoS Negl Trop Dis. 2010 Dec 21;4(12):e923.
420. Kaiser M, Bray MA, Cal M, Trunz BB, Torreele E, Brun R. Antitrypanosomal Activity of Fexinidazole, a New Oral Nitroimidazole Drug Candidate for Treatment of Sleeping Sickness. Antimicrob Agents Chemother. 2011 Dec 1;55(12):5602–8.
421. Drugs for Neglected Diseases Initiative. New drugs for neglected diseases: New hope for forgotten patients [Internet]. [cited 2014 Jan 31]. Available from: http://r4d.dfid.gov.uk/PDF/Outputs/DNDI/DNDi_Europe_brochure.pdf
422. Chesbrough H, Schwartz K. Innovating Business Models with Co-development Partnerships. Res-Technol Manag. 2007;50(1):55–9.
423. Moore GA. Crossing the chasm: marketing and selling technology products to mainstream customers. Oxford: Capstone; 2000.
424. Bray MJ, Lee JN. University revenues from technology transfer: Licensing fees vs. equity positions. J Bus Ventur. 2000 Sep;15(5–6):385–92.
425. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. J Health Econ. 2003 Mar;22(2):151–85.
426. Reinventing Life Science Startups–Therapeutics and Diagnostics | Steve Blank [Internet]. [cited 2014 Feb 6]. Available from: http://steveblank.com/2013/08/19/reinventing-life-science-startups-evidence-based-entrepreneurship/
427. Research C for DE and. Information for Consumers (Drugs) - FDA's Drug Review Process: Continued [Internet]. [cited 2014 Feb 6]. Available from: http://www.fda.gov/Drugs/ResourcesForYou/Consumers/ucm289601.htm
428. DiMasi JA, Feldman L, Seckler A, Wilson A. Trends in Risks Associated With New Drug Development: Success Rates for Investigational Drugs. Clin Pharmacol Ther. 2010 Mar;87(3):272–7.
429. Kalamas J, Pinkus G. The optimum time for drug licensing. Nat Rev Drug Discov. 2003 Sep;2(9):691–2.
430. Elvidge S. The changing shape of dealmaking [Internet]. Available from: http://www.nature.com/biopharmadealmakers/pdf/feature_dealmaking_web.pdf
431. Coller BS, Califf RM. Traversing the Valley of Death: A Guide to Assessing Prospects for Translational Success. Sci Transl Med. 2009 Dec 9;1(10):10cm9.
432. Auerswald PE, Branscomb LM. Valleys of Death and Darwinian Seas: Financing the Invention to Innovation Transition in the United States. J Technol Transf. 2003 Aug 1;28(3-4):227–39.

433. Therapeutics for Rare and Neglected Diseases [Internet]. [cited 2014 Feb 3]. Available from: http://www.ncats.nih.gov/research/rare-diseases/trnd/trnd.html
434. Danzon PM, Nicholson S, Pereira NS. Productivity in pharmaceutical–biotechnology R&D: the role of experience and alliances. J Health Econ. 2005 Mar;24(2):317–39.
435. Respira Design: Complex Requirements for a Simple Solution : Center for Social Innovation (CSI) [Internet]. [cited 2013 Dec 2]. Available from: http://csi.gsb.stanford.edu/respira-design-complex-requirements-simple-solution
436. Tsaioun K, Jacewicz M. De-risking drug discovery with ADDME -- avoiding drug development mistakes early. Altern Lab Anim ATLA. 2009 Sep;37 Suppl 1:47–55.
437. Anderson GF. Spurring New Research For Neglected Diseases. Health Aff (Millwood). 2009 Nov 1;28(6):1750–9.
438. Mackey TK, Liang BA. Global health policy coordination to address neglected tropical diseases. Trop Med Int Health. 2012 Sep;17(9):1053–6.
439. Moon S, Bermudez J, 't Hoen E. Innovation and access to medicines for neglected populations: Could a treaty address a broken pharmaceutical R&D system? PLoS Med. 2012;9(5).

7. Appendix

7.1 Virtual Screening

Table 7.1.1 - List of ZINC codes identifying all ligands from the NCI diversity set II. Where ZINC codes are repeated, there are multiple tautomers available for that ligand, all of which were included in the screening assay.

ZINC00000171	ZINC01572767	ZINC01706234	ZINC04771299	ZINC12341271	ZINC15990220
ZINC00000179	ZINC01572960	ZINC01706740	ZINC04773602	ZINC12341321	ZINC15990245
ZINC00000570	ZINC01572961	ZINC01706914	ZINC04775687	ZINC12341336	ZINC15990245
ZINC00001018	ZINC01572962	ZINC01706915	ZINC04775707	ZINC12358863	ZINC15990245
ZINC00001087	ZINC01572963	ZINC01707109	ZINC04775893	ZINC12363224	ZINC15990251
ZINC00001228	ZINC01573467	ZINC01707130	ZINC04776634	ZINC12367936	ZINC15990251
ZINC00001522	ZINC01573517	ZINC01707215	ZINC04777715	ZINC12368248	ZINC15990251
ZINC00001785	ZINC01573531	ZINC01707482	ZINC04777716	ZINC12368665	ZINC15990251
ZINC00001791	ZINC01573541	ZINC01707724	ZINC04777790	ZINC12417345	ZINC16051261
ZINC00001947	ZINC01573829	ZINC01708688	ZINC04777791	ZINC12417345	ZINC16051261
ZINC00002036	ZINC01573978	ZINC01709784	ZINC04777934	ZINC12504140	ZINC16051261
ZINC00002106	ZINC01574246	ZINC01709785	ZINC04777935	ZINC12504141	ZINC16920429
ZINC00005155	ZINC01574310	ZINC01710091	ZINC04777936	ZINC12504142	ZINC16920429
ZINC00013246	ZINC01574615	ZINC01710099	ZINC04777937	ZINC12670903	ZINC16941278
ZINC00014894	ZINC01574620	ZINC01710435	ZINC04783080	ZINC12670909	ZINC16948978
ZINC00021557	ZINC01574971	ZINC01710961	ZINC04783140	ZINC12670914	ZINC16948978
ZINC00024336	ZINC01575013	ZINC01711028	ZINC04783204	ZINC12670920	ZINC16948978
ZINC00028671	ZINC01575013	ZINC01711796	ZINC04783229	ZINC12670933	ZINC16948978
ZINC00031164	ZINC01575102	ZINC01712457	ZINC04783478	ZINC12671128	ZINC16948978
ZINC00031165	ZINC01576406	ZINC01712458	ZINC04783479	ZINC12671501	ZINC16951318
ZINC00031410	ZINC01577602	ZINC01712637	ZINC04783480	ZINC12671506	ZINC16951318
ZINC00035663	ZINC01577889	ZINC01712658	ZINC04783481	ZINC12671886	ZINC16951320
ZINC00035871	ZINC01578220	ZINC01712744	ZINC04786808	ZINC12671893	ZINC16951320
ZINC00039221	ZINC01578333	ZINC01712829	ZINC04786811	ZINC12671898	ZINC16951320
ZINC00039435	ZINC01579461	ZINC01712830	ZINC04786814	ZINC12671904	ZINC16954225
ZINC00039490	ZINC01579687	ZINC01712866	ZINC04791992	ZINC12672222	ZINC16969114
ZINC00041118	ZINC01579723	ZINC01713460	ZINC04797042	ZINC12672225	ZINC16978136
ZINC00054469	ZINC01579734	ZINC01714961	ZINC04800001	ZINC12672231	ZINC16978136
ZINC00056624	ZINC01579740	ZINC01714962	ZINC04821670	ZINC12672242	ZINC17105710
ZINC00057263	ZINC01579877	ZINC01716722	ZINC04821957	ZINC12672252	ZINC17105710
ZINC00057264	ZINC01580992	ZINC01717579	ZINC04822288	ZINC12959181	ZINC17108282
ZINC00057677	ZINC01581342	ZINC01718481	ZINC04822383	ZINC13000556	ZINC17125976
ZINC00057716	ZINC01581709	ZINC01718482	ZINC04822746	ZINC13042892	ZINC17147424
ZINC00058294	ZINC01582372	ZINC01718485	ZINC04822747	ZINC13042892	ZINC17147426
ZINC00061529	ZINC01583123	ZINC01718486	ZINC04823467	ZINC13086327	ZINC17147429
ZINC00065175	ZINC01583649	ZINC01718899	ZINC04824022	ZINC13086327	ZINC17147431

ZINC00065182	ZINC01583786	ZINC01718905	ZINC04824024	ZINC13099024	ZINC17149563
ZINC00066629	ZINC01584497	ZINC01719297	ZINC04824645	ZINC13099025	ZINC17149566
ZINC00068268	ZINC01584862	ZINC01719567	ZINC04824902	ZINC13099027	ZINC17149568
ZINC00073731	ZINC01585729	ZINC01719937	ZINC04830956	ZINC13099047	ZINC17149570
ZINC00075241	ZINC01586128	ZINC01719953	ZINC04831099	ZINC13099048	ZINC17285513
ZINC00076793	ZINC01586252	ZINC01721304	ZINC04831099	ZINC13099050	ZINC17285513
ZINC00080747	ZINC01586330	ZINC01722140	ZINC04877299	ZINC13099051	ZINC17287324
ZINC00081505	ZINC01586453	ZINC01722585	ZINC04878491	ZINC13099284	ZINC17287328
ZINC00081510	ZINC01587328	ZINC01724003	ZINC04879418	ZINC13099549	ZINC17287332
ZINC00084613	ZINC01587338	ZINC01724376	ZINC04887389	ZINC13099550	ZINC17287336
ZINC00084617	ZINC01587338	ZINC01724593	ZINC04887558	ZINC13099551	ZINC17303059
ZINC00084716	ZINC01588812	ZINC01724660	ZINC04896131	ZINC13110586	ZINC17353911
ZINC00085895	ZINC01589060	ZINC01725942	ZINC04896472	ZINC13110587	ZINC17353912
ZINC00086102	ZINC01589436	ZINC01726729	ZINC04896601	ZINC13110587	ZINC17353913
ZINC00091200	ZINC01589497	ZINC01726736	ZINC04900076	ZINC13115350	ZINC17353914
ZINC00091560	ZINC01589627	ZINC01726776	ZINC04900874	ZINC13115350	ZINC17375647
ZINC00092157	ZINC01590480	ZINC01728118	ZINC04901375	ZINC13125763	ZINC17375649
ZINC00093738	ZINC01590864	ZINC01728119	ZINC04918956	ZINC13125763	ZINC17375651
ZINC00094441	ZINC01590927	ZINC01728120	ZINC04918963	ZINC13125795	ZINC17375653
ZINC00095158	ZINC01591881	ZINC01728503	ZINC04918969	ZINC13125795	ZINC17379198
ZINC00099494	ZINC01591915	ZINC01728821	ZINC04934182	ZINC13130011	ZINC17419163
ZINC00099827	ZINC01591950	ZINC01728888	ZINC04962003	ZINC13130011	ZINC17419163
ZINC00100248	ZINC01591957	ZINC01729467	ZINC04963162	ZINC13130011	ZINC17419163
ZINC00102611	ZINC01592023	ZINC01729523	ZINC04963453	ZINC13130015	ZINC17424892
ZINC00104475	ZINC01592410	ZINC01729524	ZINC04963778	ZINC13130016	ZINC17424892
ZINC00105181	ZINC01592974	ZINC01729525	ZINC04963779	ZINC13130018	ZINC17465817
ZINC00105309	ZINC01593556	ZINC01729526	ZINC04974338	ZINC13130018	ZINC17465824
ZINC00110358	ZINC01593739	ZINC01729549	ZINC04974725	ZINC13130035	ZINC17465958
ZINC00110365	ZINC01595620	ZINC01729578	ZINC04984366	ZINC13130183	ZINC17465965
ZINC00113222	ZINC01596285	ZINC01729618	ZINC04990829	ZINC13130187	ZINC17465972
ZINC00113984	ZINC01596698	ZINC01730238	ZINC05011996	ZINC13130187	ZINC17465979
ZINC00115489	ZINC01597168	ZINC01731085	ZINC05011999	ZINC13130211	ZINC17465983
ZINC00116869	ZINC01599183	ZINC01732051	ZINC05015095	ZINC13130211	ZINC17529946
ZINC00117074	ZINC01601483	ZINC01732509	ZINC05027641	ZINC13132551	ZINC17529946
ZINC00117074	ZINC01602509	ZINC01733235	ZINC05030617	ZINC13132551	ZINC17729070
ZINC00119004	ZINC01602510	ZINC01733511	ZINC05033796	ZINC13136204	ZINC17729070
ZINC00119372	ZINC01602855	ZINC01734080	ZINC05064462	ZINC13136204	ZINC17820077
ZINC00121415	ZINC01602856	ZINC01734092	ZINC05086138	ZINC13139292	ZINC17858074
ZINC00126096	ZINC01602938	ZINC01734413	ZINC05086225	ZINC13139292	ZINC17858074
ZINC00135449	ZINC01602938	ZINC01734846	ZINC05105895	ZINC13139298	ZINC17858074
ZINC00135453	ZINC01603961	ZINC01735469	ZINC05114817	ZINC13139306	ZINC17860685
ZINC00135673	ZINC01604104	ZINC01735767	ZINC05115990	ZINC13139308	ZINC17860685
ZINC00136890	ZINC01604214	ZINC01735773	ZINC05120864	ZINC13139308	ZINC17860685
ZINC00136892	ZINC01605775	ZINC01736227	ZINC05120916	ZINC13139316	ZINC17861701

ZINC00136893	ZINC01606051	ZINC01736228	ZINC05120916	ZINC13139317	ZINC17861701
ZINC00136894	ZINC01607786	ZINC01737161	ZINC05120920	ZINC13139331	ZINC17878196
ZINC00137148	ZINC01608678	ZINC01737351	ZINC05120920	ZINC13140104	ZINC17885079
ZINC00137187	ZINC01608684	ZINC01738764	ZINC05124931	ZINC13140224	ZINC17885079
ZINC00137189	ZINC01608764	ZINC01738918	ZINC05124957	ZINC13140224	ZINC17886718
ZINC00138096	ZINC01608817	ZINC01739349	ZINC05124960	ZINC13140225	ZINC17886718
ZINC00138986	ZINC01608819	ZINC01739733	ZINC05137909	ZINC13140225	ZINC17949075
ZINC00139367	ZINC01609078	ZINC01740602	ZINC05137909	ZINC13140234	ZINC17949075
ZINC00139370	ZINC01609695	ZINC01741812	ZINC05176310	ZINC13140243	ZINC17949075
ZINC00141916	ZINC01615121	ZINC01742542	ZINC05176310	ZINC13140243	ZINC17968970
ZINC00143079	ZINC01616692	ZINC01744749	ZINC05180959	ZINC13140245	ZINC17968970
ZINC00143743	ZINC01617170	ZINC01744953	ZINC05193477	ZINC13140246	ZINC17968970
ZINC00145517	ZINC01617301	ZINC01745539	ZINC05193477	ZINC13140246	ZINC17970262
ZINC00150151	ZINC01620955	ZINC01747258	ZINC05201470	ZINC13142971	ZINC17970262
ZINC00151887	ZINC01621536	ZINC01747274	ZINC05208303	ZINC13142972	ZINC17995347
ZINC00154608	ZINC01621931	ZINC01747299	ZINC05217440	ZINC13142989	ZINC17995347
ZINC00154792	ZINC01621981	ZINC01747332	ZINC05217831	ZINC13142989	ZINC17995347
ZINC00154832	ZINC01621997	ZINC01748006	ZINC05220390	ZINC13142991	ZINC18006265
ZINC00154888	ZINC01622198	ZINC01748056	ZINC05220390	ZINC13142991	ZINC18006265
ZINC00155089	ZINC01622269	ZINC01748097	ZINC05220390	ZINC13142993	ZINC18006265
ZINC00155292	ZINC01625052	ZINC01748594	ZINC05279843	ZINC13143008	ZINC18010927
ZINC00155857	ZINC01625094	ZINC01748908	ZINC05328871	ZINC13143009	ZINC18010927
ZINC00156701	ZINC01625106	ZINC01749571	ZINC05369992	ZINC13143011	ZINC18030775
ZINC00157210	ZINC01625114	ZINC01751425	ZINC05386901	ZINC13143011	ZINC18030775
ZINC00157625	ZINC01625418	ZINC01751437	ZINC05390066	ZINC13143017	ZINC18038389
ZINC00160306	ZINC01625444	ZINC01752308	ZINC05390471	ZINC13143019	ZINC18038389
ZINC00160781	ZINC01625746	ZINC01752784	ZINC05392913	ZINC13143021	ZINC18043993
ZINC00161631	ZINC01626802	ZINC01752973	ZINC05420911	ZINC13144115	ZINC18043993
ZINC00161700	ZINC01626803	ZINC01753336	ZINC05420911	ZINC13144116	ZINC18055497
ZINC00163247	ZINC01627101	ZINC01753723	ZINC05421282	ZINC13144542	ZINC18056732
ZINC00163657	ZINC01627302	ZINC01753725	ZINC05421282	ZINC13144560	ZINC18056732
ZINC00166215	ZINC01627329	ZINC01753735	ZINC05421739	ZINC13144560	ZINC18056732
ZINC00188345	ZINC01627377	ZINC01753761	ZINC05421739	ZINC13144564	ZINC18056840
ZINC00204666	ZINC01627439	ZINC01754899	ZINC05422109	ZINC13144566	ZINC18056840
ZINC00214590	ZINC01628211	ZINC01755448	ZINC05422109	ZINC13144568	ZINC18057104
ZINC00217297	ZINC01629421	ZINC01755627	ZINC05424472	ZINC13144613	ZINC18057104
ZINC00225990	ZINC01629569	ZINC01757986	ZINC05424474	ZINC13144614	ZINC18057104
ZINC00226959	ZINC01629851	ZINC01761347	ZINC05424691	ZINC13144616	ZINC18057933
ZINC00227103	ZINC01632739	ZINC01761361	ZINC05424691	ZINC13144618	ZINC18057933
ZINC00235625	ZINC01633473	ZINC01761642	ZINC05430677	ZINC13144621	ZINC18061739
ZINC00235981	ZINC01633669	ZINC01761875	ZINC05434498	ZINC13145027	ZINC18067025
ZINC00247785	ZINC01635676	ZINC01761888	ZINC05439444	ZINC13145028	ZINC18067025
ZINC00258800	ZINC01635813	ZINC01764014	ZINC05439444	ZINC13145029	ZINC18067025
ZINC00270777	ZINC01635835	ZINC01820901	ZINC05452603	ZINC13145030	ZINC18068040

ZINC00274634	ZINC01635869	ZINC01820901	ZINC05452603	ZINC13146918	ZINC18068040
ZINC00274689	ZINC01637135	ZINC01834023	ZINC05459826	ZINC13146942	ZINC18068788
ZINC00282116	ZINC01637139	ZINC01843029	ZINC05462670	ZINC13146950	ZINC18083898
ZINC00283058	ZINC01637206	ZINC01843029	ZINC05462674	ZINC13146950	ZINC18083898
ZINC00292958	ZINC01637653	ZINC01843030	ZINC05479118	ZINC13146954	ZINC18085106
ZINC00293214	ZINC01638088	ZINC01843030	ZINC05479451	ZINC13146954	ZINC18085106
ZINC00304325	ZINC01638353	ZINC01843071	ZINC05487838	ZINC13146957	ZINC18085106
ZINC00306008	ZINC01638368	ZINC01845586	ZINC05487838	ZINC13152217	ZINC18098114
ZINC00323854	ZINC01639633	ZINC01845598	ZINC05492794	ZINC13152217	ZINC18098114
ZINC00329885	ZINC01639634	ZINC01846592	ZINC05493068	ZINC13152219	ZINC18098743
ZINC00330473	ZINC01640145	ZINC01846592	ZINC05493736	ZINC13152221	ZINC18098743
ZINC00330473	ZINC01640193	ZINC01848197	ZINC05498636	ZINC13152223	ZINC18098743
ZINC00331547	ZINC01640205	ZINC01848198	ZINC05498636	ZINC13152225	ZINC18098743
ZINC00331745	ZINC01640219	ZINC01855333	ZINC05499702	ZINC13152226	ZINC18107657
ZINC00332365	ZINC01641160	ZINC01856534	ZINC05499876	ZINC13152237	ZINC18117772
ZINC00337138	ZINC01641244	ZINC01865671	ZINC05499876	ZINC13152238	ZINC18121761
ZINC00350319	ZINC01641376	ZINC01867048	ZINC05502368	ZINC13152238	ZINC18124911
ZINC00369656	ZINC01641466	ZINC01868209	ZINC05502368	ZINC13152247	ZINC18124911
ZINC00383488	ZINC01641644	ZINC01871103	ZINC05513820	ZINC13152247	ZINC18126453
ZINC00383489	ZINC01644016	ZINC01871223	ZINC05513820	ZINC13152248	ZINC18126453
ZINC00387464	ZINC01644254	ZINC01871349	ZINC05518987	ZINC13152249	ZINC18141142
ZINC00389747	ZINC01645454	ZINC01872881	ZINC05536814	ZINC13152277	ZINC18141294
ZINC00389762	ZINC01646278	ZINC01873142	ZINC05541927	ZINC13152277	ZINC18141294
ZINC00392897	ZINC01646387	ZINC02011568	ZINC05541929	ZINC13152278	ZINC18141294
ZINC00393546	ZINC01646690	ZINC02015922	ZINC05541931	ZINC13152279	ZINC18141294
ZINC00393598	ZINC01648249	ZINC02018030	ZINC05541933	ZINC13152282	ZINC18141294
ZINC00393599	ZINC01649010	ZINC02018030	ZINC05550368	ZINC13152284	ZINC18142874
ZINC00393651	ZINC01649013	ZINC02029995	ZINC05550368	ZINC13152291	ZINC18153859
ZINC00393674	ZINC01649032	ZINC02030136	ZINC05550663	ZINC13152292	ZINC18153859
ZINC00393856	ZINC01652261	ZINC02033748	ZINC05550663	ZINC13152294	ZINC18154478
ZINC00393885	ZINC01652969	ZINC02034999	ZINC05550666	ZINC13152611	ZINC18154478
ZINC00394316	ZINC01652984	ZINC02035165	ZINC05550666	ZINC13152612	ZINC18154478
ZINC00394330	ZINC01652986	ZINC02036484	ZINC05551091	ZINC13152612	ZINC18156788
ZINC00395036	ZINC01653222	ZINC02036556	ZINC05551091	ZINC13152613	ZINC18156788
ZINC00395364	ZINC01655764	ZINC02038558	ZINC05551091	ZINC13152613	ZINC18157167
ZINC00395576	ZINC01655914	ZINC02042220	ZINC05551760	ZINC13152614	ZINC18163183
ZINC00395576	ZINC01656025	ZINC02042648	ZINC05551760	ZINC13152614	ZINC18163421
ZINC00395872	ZINC01656026	ZINC02042811	ZINC05551760	ZINC13152615	ZINC18163421
ZINC00397002	ZINC01656027	ZINC02042840	ZINC05552274	ZINC13152615	ZINC18163421
ZINC00397404	ZINC01656028	ZINC02042931	ZINC05552274	ZINC13152616	ZINC18166466
ZINC00397978	ZINC01656323	ZINC02043260	ZINC05554415	ZINC13154271	ZINC18166466
ZINC00398060	ZINC01657060	ZINC02045340	ZINC05554417	ZINC13154273	ZINC18166590
ZINC00398332	ZINC01659930	ZINC02046452	ZINC05575116	ZINC13154282	ZINC18168901
ZINC00398390	ZINC01661542	ZINC02046726	ZINC05576272	ZINC13154298	ZINC18169477

ZINC00398392	ZINC01661545	ZINC02048098	ZINC05576272	ZINC13154298	ZINC18169477
ZINC00398470	ZINC01662442	ZINC02048214	ZINC05580556	ZINC13154298	ZINC18169477
ZINC00399102	ZINC01663950	ZINC02048272	ZINC05580600	ZINC13154300	ZINC18169477
ZINC00399281	ZINC01663959	ZINC02048695	ZINC05580600	ZINC13154302	ZINC18169477
ZINC00399404	ZINC01665042	ZINC02154763	ZINC05580712	ZINC13154304	ZINC18173699
ZINC00399512	ZINC01665187	ZINC02169824	ZINC05580712	ZINC13154305	ZINC18178799
ZINC00400127	ZINC01665192	ZINC02476372	ZINC05581222	ZINC13154306	ZINC18183236
ZINC00400410	ZINC01665194	ZINC02476372	ZINC05581222	ZINC13154307	ZINC18189380
ZINC00400541	ZINC01665334	ZINC02476372	ZINC05581974	ZINC13154313	ZINC18189380
ZINC00401809	ZINC01665703	ZINC02515938	ZINC05581974	ZINC13154314	ZINC18189380
ZINC00404349	ZINC01665708	ZINC02670038	ZINC05594334	ZINC13154315	ZINC18192390
ZINC00404427	ZINC01665801	ZINC02952832	ZINC05594334	ZINC13154315	ZINC18192390
ZINC00404427	ZINC01665928	ZINC03100631	ZINC05640172	ZINC13154317	ZINC18193021
ZINC00404427	ZINC01665940	ZINC03157611	ZINC05640410	ZINC13154317	ZINC18193844
ZINC00404472	ZINC01666192	ZINC03258337	ZINC05640410	ZINC13154319	ZINC18193844
ZINC00407984	ZINC01666568	ZINC03310974	ZINC05640600	ZINC13154320	ZINC18193844
ZINC00408303	ZINC01666997	ZINC03647770	ZINC05640620	ZINC13154321	ZINC18207613
ZINC00408598	ZINC01668114	ZINC03830315	ZINC05641037	ZINC13154324	ZINC18211623
ZINC00409811	ZINC01668253	ZINC03830316	ZINC05647206	ZINC13154325	ZINC18212500
ZINC00409897	ZINC01668429	ZINC03831067	ZINC05647206	ZINC13154326	ZINC18213583
ZINC00410187	ZINC01668491	ZINC03831068	ZINC05647206	ZINC13154328	ZINC18217586
ZINC00410188	ZINC01668601	ZINC03832421	ZINC05662861	ZINC13154328	ZINC18219562
ZINC00410244	ZINC01668706	ZINC03832421	ZINC05664697	ZINC13154329	ZINC18219564
ZINC00410252	ZINC01668896	ZINC03845141	ZINC05664697	ZINC13154329	ZINC18227443
ZINC00410253	ZINC01669376	ZINC03850501	ZINC05665089	ZINC13207617	ZINC18825330
ZINC00410269	ZINC01669572	ZINC03860856	ZINC05709639	ZINC13207617	ZINC18847035
ZINC00410269	ZINC01669828	ZINC03860920	ZINC05723045	ZINC13208966	ZINC19205460
ZINC00410312	ZINC01670127	ZINC03860960	ZINC05723051	ZINC13281671	ZINC19230120
ZINC00433226	ZINC01670214	ZINC03860960	ZINC05732213	ZINC13282591	ZINC19322683
ZINC00434844	ZINC01670291	ZINC03875548	ZINC05782788	ZINC13284902	ZINC19325788
ZINC00442499	ZINC01670341	ZINC03878829	ZINC05785035	ZINC13451125	ZINC19325788
ZINC00444253	ZINC01670393	ZINC03881813	ZINC05785035	ZINC13522091	ZINC19325788
ZINC00475774	ZINC01670696	ZINC03881918	ZINC05810714	ZINC13542620	ZINC19325791
ZINC00513614	ZINC01670877	ZINC03884497	ZINC05810741	ZINC13542620	ZINC19325791
ZINC00538110	ZINC01671311	ZINC03884497	ZINC05824155	ZINC13542876	ZINC19325794
ZINC00563913	ZINC01671321	ZINC03898665	ZINC05834946	ZINC13542876	ZINC19325794
ZINC00600322	ZINC01671640	ZINC03898710	ZINC05834946	ZINC13597216	ZINC19325794
ZINC00601312	ZINC01671654	ZINC03898822	ZINC05844151	ZINC13597216	ZINC19325794
ZINC00607731	ZINC01671866	ZINC03947435	ZINC05979393	ZINC13597218	ZINC19362650
ZINC00608128	ZINC01671868	ZINC03953394	ZINC05993141	ZINC13597218	ZINC19362651
ZINC00608205	ZINC01671896	ZINC03953472	ZINC05997896	ZINC13597219	ZINC19366110
ZINC00611334	ZINC01672033	ZINC03953805	ZINC06092501	ZINC13597219	ZINC19366557
ZINC00611335	ZINC01672062	ZINC03953810	ZINC06092501	ZINC13597222	ZINC19456377
ZINC00615883	ZINC01672214	ZINC03953920	ZINC06092503	ZINC13597222	ZINC19456379

ZINC00625965	ZINC01672597	ZINC03954247	ZINC06092503	ZINC13597230	ZINC19632726
ZINC00639634	ZINC01672599	ZINC03954311	ZINC06095736	ZINC13597230	ZINC19701771
ZINC00640726	ZINC01673377	ZINC03954351	ZINC06095736	ZINC13597232	ZINC19735376
ZINC00640777	ZINC01673464	ZINC03954397	ZINC06188101	ZINC13597232	ZINC19735376
ZINC00641093	ZINC01674390	ZINC03954507	ZINC06343280	ZINC13597237	ZINC19794764
ZINC00641146	ZINC01674410	ZINC03954520	ZINC06381592	ZINC13597237	ZINC19794765
ZINC00642588	ZINC01674943	ZINC03954552	ZINC06511573	ZINC13597273	ZINC19799271
ZINC00728291	ZINC01675432	ZINC03954626	ZINC06511573	ZINC13597273	ZINC19909409
ZINC00810178	ZINC01675804	ZINC03958471	ZINC06511656	ZINC13597276	ZINC19923695
ZINC00815686	ZINC01675857	ZINC04006098	ZINC06511656	ZINC13597276	ZINC19923697
ZINC00838233	ZINC01675858	ZINC04015433	ZINC06511739	ZINC13597278	ZINC20028475
ZINC00849668	ZINC01675990	ZINC04015715	ZINC06511739	ZINC13597278	ZINC20421310
ZINC00849668	ZINC01676023	ZINC04027061	ZINC06511825	ZINC13597281	ZINC21673561
ZINC00896836	ZINC01676138	ZINC04027061	ZINC06511825	ZINC13597281	ZINC22267016
ZINC00900951	ZINC01676139	ZINC04095812	ZINC06513204	ZINC13597298	ZINC22586514
ZINC00967339	ZINC01676213	ZINC04127920	ZINC06513204	ZINC13597298	ZINC22910159
ZINC00968360	ZINC01676257	ZINC04164657	ZINC06513918	ZINC13597334	ZINC22910179
ZINC00977107	ZINC01676311	ZINC04212654	ZINC06513918	ZINC13597335	ZINC22910859
ZINC00984004	ZINC01677276	ZINC04213833	ZINC06520320	ZINC13597335	ZINC22910942
ZINC00990239	ZINC01677554	ZINC04213833	ZINC06576021	ZINC13597337	ZINC22911751
ZINC01036792	ZINC01677761	ZINC04213833	ZINC06576021	ZINC13597337	ZINC22911751
ZINC01036880	ZINC01677892	ZINC04214344	ZINC06576323	ZINC13597346	ZINC22911959
ZINC01045051	ZINC01679310	ZINC04214836	ZINC06576323	ZINC13597348	ZINC22912001
ZINC01045051	ZINC01679777	ZINC04217305	ZINC06576501	ZINC13597350	ZINC22912013
ZINC01045089	ZINC01680284	ZINC04217587	ZINC06576501	ZINC13597354	ZINC22912016
ZINC01045090	ZINC01680565	ZINC04237573	ZINC06857068	ZINC13597354	ZINC22918614
ZINC01045530	ZINC01681130	ZINC04347718	ZINC06857068	ZINC13597363	ZINC23118752
ZINC01045531	ZINC01681557	ZINC04353020	ZINC07996353	ZINC13597363	ZINC23118754
ZINC01069082	ZINC01682029	ZINC04365786	ZINC08553055	ZINC13597368	ZINC23118770
ZINC01081577	ZINC01682117	ZINC04366202	ZINC08553055	ZINC13597368	ZINC23118772
ZINC01082213	ZINC01682520	ZINC04366261	ZINC08560067	ZINC13597372	ZINC23211906
ZINC01082213	ZINC01682534	ZINC04366520	ZINC08580873	ZINC13597383	ZINC23211908
ZINC01163259	ZINC01682798	ZINC04366893	ZINC08581164	ZINC13597383	ZINC24334040
ZINC01196937	ZINC01683295	ZINC04367141	ZINC08581164	ZINC13597390	ZINC25783872
ZINC01323776	ZINC01683317	ZINC04376856	ZINC08581166	ZINC13597390	ZINC26730911
ZINC01481770	ZINC01683345	ZINC04403652	ZINC08581166	ZINC13597410	ZINC26780933
ZINC01481981	ZINC01683531	ZINC04403653	ZINC08581317	ZINC13597410	ZINC26780935
ZINC01530795	ZINC01683553	ZINC04409180	ZINC08581317	ZINC13597416	ZINC26780937
ZINC01530836	ZINC01683648	ZINC04416197	ZINC08581421	ZINC13597416	ZINC26780939
ZINC01532548	ZINC01685254	ZINC04416613	ZINC08581460	ZINC13597418	ZINC29589543
ZINC01555262	ZINC01685697	ZINC04428312	ZINC08585953	ZINC13597418	ZINC29589553
ZINC01555547	ZINC01685876	ZINC04428312	ZINC08603298	ZINC13597424	ZINC29589777
ZINC01555676	ZINC01685881	ZINC04428581	ZINC08603298	ZINC13597426	ZINC29589785
ZINC01555979	ZINC01685966	ZINC04428843	ZINC08603298	ZINC13597428	ZINC29589797

ZINC01556940	ZINC01686251	ZINC04428858	ZINC08613703	ZINC13597428	ZINC29589800
ZINC01558226	ZINC01686363	ZINC04428999	ZINC08613703	ZINC13597430	ZINC29589816
ZINC01559756	ZINC01686467	ZINC04429620	ZINC08615383	ZINC13597430	ZINC29589818
ZINC01560078	ZINC01686683	ZINC04430482	ZINC08615531	ZINC13597432	ZINC29589820
ZINC01560636	ZINC01686969	ZINC04430753	ZINC08615531	ZINC13597432	ZINC29589822
ZINC01561127	ZINC01687247	ZINC04473135	ZINC08615799	ZINC13597717	ZINC29589824
ZINC01561462	ZINC01687845	ZINC04522231	ZINC08616208	ZINC13597717	ZINC29589826
ZINC01561576	ZINC01687910	ZINC04522477	ZINC08616208	ZINC13597721	ZINC29589828
ZINC01561637	ZINC01688151	ZINC04531729	ZINC08617619	ZINC13597721	ZINC29589829
ZINC01561931	ZINC01688347	ZINC04536747	ZINC08617619	ZINC13597728	ZINC29589831
ZINC01562022	ZINC01688628	ZINC04536750	ZINC08618545	ZINC13597730	ZINC29589833
ZINC01562304	ZINC01688646	ZINC04536752	ZINC08618968	ZINC13597730	ZINC29589835
ZINC01562340	ZINC01688662	ZINC04537227	ZINC08622475	ZINC13597732	ZINC29589837
ZINC01562352	ZINC01688752	ZINC04537229	ZINC08622475	ZINC13597732	ZINC29589855
ZINC01562357	ZINC01689447	ZINC04537816	ZINC08627485	ZINC13597738	ZINC29589859
ZINC01562359	ZINC01689932	ZINC04538450	ZINC08627502	ZINC13597741	ZINC29589876
ZINC01562368	ZINC01690169	ZINC04547321	ZINC08637986	ZINC13597741	ZINC29589877
ZINC01562588	ZINC01690171	ZINC04557220	ZINC08648354	ZINC13597743	ZINC29589879
ZINC01563104	ZINC01690194	ZINC04557220	ZINC08648354	ZINC13597743	ZINC29589881
ZINC01563302	ZINC01690208	ZINC04558887	ZINC08649723	ZINC13597746	ZINC29589883
ZINC01563324	ZINC01690699	ZINC04578913	ZINC08649723	ZINC13597746	ZINC29589885
ZINC01563325	ZINC01690722	ZINC04628938	ZINC08649758	ZINC13597758	ZINC29589888
ZINC01563326	ZINC01691075	ZINC04683414	ZINC08649758	ZINC13597762	ZINC29589897
ZINC01563327	ZINC01691548	ZINC04691860	ZINC08652230	ZINC13597762	ZINC29589899
ZINC01563973	ZINC01691551	ZINC04692015	ZINC08652230	ZINC13597764	ZINC29589901
ZINC01563997	ZINC01691651	ZINC04692016	ZINC08655995	ZINC13597764	ZINC29589906
ZINC01563998	ZINC01691782	ZINC04701260	ZINC08655995	ZINC13597767	ZINC29589907
ZINC01564760	ZINC01691943	ZINC04701343	ZINC08660420	ZINC13597767	ZINC29589912
ZINC01566093	ZINC01691989	ZINC04705802	ZINC08672844	ZINC13597771	ZINC29589917
ZINC01566193	ZINC01692470	ZINC04706168	ZINC08672844	ZINC13597771	ZINC29589923
ZINC01566453	ZINC01693279	ZINC04707417	ZINC08772958	ZINC13597814	ZINC29589924
ZINC01566685	ZINC01694053	ZINC04707457	ZINC08772958	ZINC13597816	ZINC29590257
ZINC01566905	ZINC01694371	ZINC04707664	ZINC09211296	ZINC13597818	ZINC29590259
ZINC01566917	ZINC01695372	ZINC04707767	ZINC09211296	ZINC13634510	ZINC29590262
ZINC01567155	ZINC01695559	ZINC04707792	ZINC09230252	ZINC13634510	ZINC29590263
ZINC01567413	ZINC01696555	ZINC04707806	ZINC11525688	ZINC13634517	ZINC29590275
ZINC01567421	ZINC01697077	ZINC04715161	ZINC11525688	ZINC13634517	ZINC29590277
ZINC01567747	ZINC01697110	ZINC04720969	ZINC11525688	ZINC13634519	ZINC29590287
ZINC01568035	ZINC01697295	ZINC04720972	ZINC11535692	ZINC13634519	ZINC29590289
ZINC01568661	ZINC01697912	ZINC04726283	ZINC11535830	ZINC14717925	ZINC29590292
ZINC01568793	ZINC01699287	ZINC04726283	ZINC11535850	ZINC15881824	ZINC29590296
ZINC01568966	ZINC01699937	ZINC04726283	ZINC11616854	ZINC15889046	ZINC29590298
ZINC01569015	ZINC01700219	ZINC04726284	ZINC11616855	ZINC15889046	ZINC29590300
ZINC01569236	ZINC01700953	ZINC04726284	ZINC11616856	ZINC15889046	ZINC29590302

ZINC01569237	ZINC01701112	ZINC04726284	ZINC11616857	ZINC15894745	ZINC29590304
ZINC01569416	ZINC01701460	ZINC04726655	ZINC11677161	ZINC15924502	ZINC29590306
ZINC01569482	ZINC01701513	ZINC04726869	ZINC11677168	ZINC15924502	ZINC29590308
ZINC01569495	ZINC01702739	ZINC04742623	ZINC11677172	ZINC15924534	ZINC29590315
ZINC01569498	ZINC01702835	ZINC04743527	ZINC11677178	ZINC15924536	ZINC29590323
ZINC01569586	ZINC01703050	ZINC04744282	ZINC11681161	ZINC15924536	
ZINC01570210	ZINC01703105	ZINC04747919	ZINC11681164	ZINC15924541	
ZINC01570216	ZINC01703109	ZINC04748128	ZINC11681166	ZINC15924541	
ZINC01570386	ZINC01703286	ZINC04748141	ZINC11681170	ZINC15952559	
ZINC01570468	ZINC01705443	ZINC04758411	ZINC12153441	ZINC15952559	
ZINC01571105	ZINC01705919	ZINC04760065	ZINC12153442	ZINC15952559	
ZINC01571135	ZINC01705925	ZINC04769176	ZINC12340155	ZINC15952610	
ZINC01571136	ZINC01706083	ZINC04769447	ZINC12340238	ZINC15952610	
ZINC01572056	ZINC01706126	ZINC04769448	ZINC12340238	ZINC15990220	
ZINC01572309	ZINC01706223	ZINC04769600	ZINC12341271	ZINC15990220	

Table 7.1.2 - ZINC codes identifying all the ligands screened in the Ligand Substructure library.

ZINC00000066	ZINC00025057	ZINC00031967	ZINC01556388	ZINC00006373	ZINC02629278
ZINC00000073	ZINC00025162	ZINC00140815	ZINC01556581	ZINC00020000	ZINC02636757
ZINC00000079	ZINC00025502	ZINC00140817	ZINC01556707	ZINC00020001	ZINC02637030
ZINC00000135	ZINC00025667	ZINC00140865	ZINC01556732	ZINC00022012	ZINC02641146
ZINC00000136	ZINC00025668	ZINC00140867	ZINC01560152	ZINC00088505	ZINC02840177
ZINC00000171	ZINC00025672	ZINC00159509	ZINC01562219	ZINC00120348	ZINC02922103
ZINC00000238	ZINC00026217	ZINC00409994	ZINC01562556	ZINC00132664	ZINC02922160
ZINC00000344	ZINC00026220	ZINC00895711	ZINC01564734	ZINC00156034	ZINC02922173
ZINC00000351	ZINC00027448	ZINC00901485	ZINC01567073	ZINC00156040	ZINC02922204
ZINC00000353	ZINC00027657	ZINC01529162	ZINC01568417	ZINC00156683	ZINC02922222
ZINC00000376	ZINC00028212	ZINC01529165	ZINC01580781	ZINC00160722	ZINC03074214
ZINC00000419	ZINC00031377	ZINC01529166	ZINC01581307	ZINC00188454	ZINC03074417
ZINC00000425	ZINC00031378	ZINC01529167	ZINC01583273	ZINC00344019	ZINC03078171
ZINC00000437	ZINC00031379	ZINC01529168	ZINC01583794	ZINC00374875	ZINC03078173
ZINC00000450	ZINC00031380	ZINC01529331	ZINC01587239	ZINC00391960	ZINC03091289
ZINC00000451	ZINC00032819	ZINC01529332	ZINC01587728	ZINC00402724	ZINC03102125
ZINC00000458	ZINC00032820	ZINC01529333	ZINC01588564	ZINC00404775	ZINC03102127
ZINC00000665	ZINC00033480	ZINC01529334	ZINC01588565	ZINC00405901	ZINC03138946
ZINC00000745	ZINC00033887	ZINC01529422	ZINC01588566	ZINC00517939	ZINC03159486
ZINC00000863	ZINC00035526	ZINC01529453	ZINC01590332	ZINC00600877	ZINC03160646
ZINC00000964	ZINC00035527	ZINC01529613	ZINC01590993	ZINC00874515	ZINC03167464
ZINC00001043	ZINC00035528	ZINC01529656	ZINC01591673	ZINC00895103	ZINC03194831
ZINC00001063	ZINC00035529	ZINC01529657	ZINC01591674	ZINC00901628	ZINC03194832
ZINC00001253	ZINC00035530	ZINC01529662	ZINC01591675	ZINC00901878	ZINC03201336
ZINC00001332	ZINC00035531	ZINC01529702	ZINC01591676	ZINC00938080	ZINC03221617
ZINC00001368	ZINC00035532	ZINC01529703	ZINC01591695	ZINC01000303	ZINC03260391
ZINC00001457	ZINC00035533	ZINC01529717	ZINC01592347	ZINC01481961	ZINC03273581
ZINC00001484	ZINC00036467	ZINC01529730	ZINC01592549	ZINC01484069	ZINC03273629
ZINC00001505	ZINC00037880	ZINC01529739	ZINC01593340	ZINC01530328	ZINC03280808
ZINC00001510	ZINC00039033	ZINC01529742	ZINC01594163	ZINC01530514	ZINC03295099
ZINC00001714	ZINC00039093	ZINC01529771	ZINC01594165	ZINC01531686	ZINC03295466
ZINC00001761	ZINC00039097	ZINC01529785	ZINC01594443	ZINC01532439	ZINC03296892
ZINC00002053	ZINC00039105	ZINC01529791	ZINC01594996	ZINC01532464	ZINC03297153
ZINC00002158	ZINC00039860	ZINC01529813	ZINC01594997	ZINC01532909	ZINC03313360
ZINC00002159	ZINC00039896	ZINC01529814	ZINC01594998	ZINC01542081	ZINC03313499
ZINC00002219	ZINC00039936	ZINC01529815	ZINC01594999	ZINC01560901	ZINC03314644
ZINC00002270	ZINC00039940	ZINC01529816	ZINC01595997	ZINC01560905	ZINC03314942
ZINC00002617	ZINC00040153	ZINC01529821	ZINC01596674	ZINC01564877	ZINC03314949
ZINC00003758	ZINC00040169	ZINC01529835	ZINC01596916	ZINC01568160	ZINC03315231
ZINC00003821	ZINC00040170	ZINC01529836	ZINC01597148	ZINC01576233	ZINC03365068
ZINC00003901	ZINC00040750	ZINC01529837	ZINC01597149	ZINC01576240	ZINC03399313
ZINC00003950	ZINC00040753	ZINC01529842	ZINC01597150	ZINC01576312	ZINC03412817
ZINC00004023	ZINC00040762	ZINC01529843	ZINC01597151	ZINC01584749	ZINC03417739

ZINC00004235	ZINC00040788	ZINC01529844	ZINC01597152	ZINC01594604	ZINC03426346
ZINC00004253	ZINC00040789	ZINC01529846	ZINC01597642	ZINC01594663	ZINC03450331
ZINC00004302	ZINC00041059	ZINC01529847	ZINC01598781	ZINC01604030	ZINC03645182
ZINC00004477	ZINC00041563	ZINC01529876	ZINC01600987	ZINC01605652	ZINC03776577
ZINC00004489	ZINC00041578	ZINC01529879	ZINC01601594	ZINC01607133	ZINC03795528
ZINC00004512	ZINC00041974	ZINC01529880	ZINC01601632	ZINC01608082	ZINC03798304
ZINC00004564	ZINC00042934	ZINC01529882	ZINC01602330	ZINC01609614	ZINC03812004
ZINC00004565	ZINC00046810	ZINC01529884	ZINC01603125	ZINC01615672	ZINC03812006
ZINC00004611	ZINC00046935	ZINC01529894	ZINC01603935	ZINC01616666	ZINC03825190
ZINC00004695	ZINC00047136	ZINC01529895	ZINC01604873	ZINC01618145	ZINC03851725
ZINC00004775	ZINC00047505	ZINC01529896	ZINC01605141	ZINC01618870	ZINC03870147
ZINC00004959	ZINC00049988	ZINC01529897	ZINC01606708	ZINC01626244	ZINC03950412
ZINC00005057	ZINC00050168	ZINC01529904	ZINC01609019	ZINC01648586	ZINC03974740
ZINC00005081	ZINC00051582	ZINC01529910	ZINC01610563	ZINC01670229	ZINC03982716
ZINC00005285	ZINC00051583	ZINC01529911	ZINC01610564	ZINC01676922	ZINC03994526
ZINC00005618	ZINC00051802	ZINC01529921	ZINC01613865	ZINC01677514	ZINC04005102
ZINC00005619	ZINC00053379	ZINC01529927	ZINC01613871	ZINC01677635	ZINC04015559
ZINC00005637	ZINC00054575	ZINC01529966	ZINC01613876	ZINC01677745	ZINC04017060
ZINC00005744	ZINC00054811	ZINC01530017	ZINC01613879	ZINC01678285	ZINC04091747
ZINC00005980	ZINC00054812	ZINC01530031	ZINC01616291	ZINC01680241	ZINC04095553
ZINC00006044	ZINC00056325	ZINC01530142	ZINC01616296	ZINC01685560	ZINC04097222
ZINC00006228	ZINC00056350	ZINC01530145	ZINC01616299	ZINC01687472	ZINC04097230
ZINC00006508	ZINC00056426	ZINC01530146	ZINC01616305	ZINC01687705	ZINC04115756
ZINC00006532	ZINC00056610	ZINC01530152	ZINC01616311	ZINC01690366	ZINC04141542
ZINC00006614	ZINC00056649	ZINC01530154	ZINC01616324	ZINC01697988	ZINC04212164
ZINC00007255	ZINC00056665	ZINC01530214	ZINC01616326	ZINC01700276	ZINC04218962
ZINC00007502	ZINC00056702	ZINC01530215	ZINC01616847	ZINC01701174	ZINC04285766
ZINC00008065	ZINC00056703	ZINC01530216	ZINC01617401	ZINC01708924	ZINC04286556
ZINC00008288	ZINC00056707	ZINC01530217	ZINC01618272	ZINC01708940	ZINC04286569
ZINC00008809	ZINC00056719	ZINC01530236	ZINC01618273	ZINC01714305	ZINC04286577
ZINC00008846	ZINC00056733	ZINC01530238	ZINC01619910	ZINC01715387	ZINC04286584
ZINC00009162	ZINC00056779	ZINC01530239	ZINC01620030	ZINC01727442	ZINC04286590
ZINC00009698	ZINC00056780	ZINC01530250	ZINC01620619	ZINC01727446	ZINC04286600
ZINC00009699	ZINC00056784	ZINC01530252	ZINC01620621	ZINC01728070	ZINC04286611
ZINC00009700	ZINC00056785	ZINC01530288	ZINC01621156	ZINC01737910	ZINC04286619
ZINC00009701	ZINC00056786	ZINC01530302	ZINC01621874	ZINC01737911	ZINC04286623
ZINC00009702	ZINC00056787	ZINC01530327	ZINC01621895	ZINC01742240	ZINC04286630
ZINC00010512	ZINC00056818	ZINC01530374	ZINC01624118	ZINC01759522	ZINC04286641
ZINC00010937	ZINC00056834	ZINC01530375	ZINC01624131	ZINC01856861	ZINC04286652
ZINC00011837	ZINC00057063	ZINC01530376	ZINC01624499	ZINC01888747	ZINC04286663
ZINC00011953	ZINC00057124	ZINC01530549	ZINC01624526	ZINC01912288	ZINC04286672
ZINC00012205	ZINC00057125	ZINC01530551	ZINC01624527	ZINC02012354	ZINC04286694
ZINC00012206	ZINC00057160	ZINC01530552	ZINC01624528	ZINC02012356	ZINC04286709
ZINC00012207	ZINC00057174	ZINC01532262	ZINC01624582	ZINC02020209	ZINC04286722

ZINC00012209	ZINC00057180	ZINC01532517	ZINC01624583	ZINC02134706	ZINC04294964
ZINC00012210	ZINC00057182	ZINC01532519	ZINC01624584	ZINC02169490	ZINC04336688
ZINC00012215	ZINC00057187	ZINC01532521	ZINC01624585	ZINC02242978	ZINC04343978
ZINC00012216	ZINC00057191	ZINC01532540	ZINC01624619	ZINC02389623	ZINC04517289
ZINC00012343	ZINC00057232	ZINC01532553	ZINC01627309	ZINC02389704	ZINC04521590
ZINC00012425	ZINC00057259	ZINC01532558	ZINC01627834	ZINC02391019	ZINC04533814
ZINC00012426	ZINC00057265	ZINC01532568	ZINC01628282	ZINC02392274	ZINC04536644
ZINC00013840	ZINC00057268	ZINC01532569	ZINC01628572	ZINC02392314	ZINC04543577
ZINC00014545	ZINC00057272	ZINC01532570	ZINC01628574	ZINC02506775	ZINC04760162
ZINC00017142	ZINC00057276	ZINC01532571	ZINC01630373	ZINC02516985	ZINC04760877
ZINC00017145	ZINC00057285	ZINC01532578	ZINC01632503	ZINC02517036	ZINC04761950
ZINC00017591	ZINC00057286	ZINC01532584	ZINC01632504	ZINC02530801	ZINC04761954
ZINC00017592	ZINC00057289	ZINC01532613	ZINC01632505	ZINC02540543	ZINC04761975
ZINC00017654	ZINC00057290	ZINC01532617	ZINC01632506	ZINC02555050	ZINC04763802
ZINC00018468	ZINC00057296	ZINC01532646	ZINC01632507	ZINC02555074	ZINC04771016
ZINC00018469	ZINC00057301	ZINC01532647	ZINC01632508	ZINC02555075	ZINC04782844
ZINC00018560	ZINC00057303	ZINC01532648	ZINC01632509	ZINC02555076	ZINC04817272
ZINC00018570	ZINC00057318	ZINC01532649	ZINC01632510	ZINC02555094	ZINC04824842
ZINC00019686	ZINC00057330	ZINC01532655	ZINC01632620	ZINC02555095	ZINC04829022
ZINC00019968	ZINC00057331	ZINC01532684	ZINC01636511	ZINC02555097	ZINC04992328
ZINC00020543	ZINC00057340	ZINC01532715	ZINC01636512	ZINC02555102	ZINC04992567
ZINC00020956	ZINC00057341	ZINC01532721	ZINC01637108	ZINC02556509	ZINC05018126
ZINC00020957	ZINC00057371	ZINC01532725	ZINC01637220	ZINC02556684	ZINC05053838
ZINC00021027	ZINC00057376	ZINC01532726	ZINC01637452	ZINC02560007	ZINC05054972
ZINC00021028	ZINC00057377	ZINC01532759	ZINC01640849	ZINC02560008	ZINC05085332
ZINC00021247	ZINC00057381	ZINC01532776	ZINC01641636	ZINC02560349	ZINC05112434
ZINC00021950	ZINC00057393	ZINC01532789	ZINC01645659	ZINC02561115	ZINC05116006
ZINC00021951	ZINC00057394	ZINC01532800	ZINC01646394	ZINC02562507	ZINC05131962
ZINC00022805	ZINC00057396	ZINC01532808	ZINC01646626	ZINC02567344	ZINC05131963
ZINC00023158	ZINC00057398	ZINC01532854	ZINC01646924	ZINC02575124	ZINC05134904
ZINC00024461	ZINC00057404	ZINC01532864	ZINC01646927	ZINC02583053	ZINC05167476
ZINC00024462	ZINC00057405	ZINC01532865	ZINC01646928	ZINC02583054	ZINC05187268
ZINC00024463	ZINC00057411	ZINC01532883	ZINC01653793	ZINC02583268	ZINC05215425
ZINC00024464	ZINC00057415	ZINC01542524	ZINC01653813	ZINC02583269	ZINC05220648
ZINC00024552	ZINC00057416	ZINC01542526	ZINC01656089	ZINC02583274	ZINC05246877
ZINC00024609	ZINC00057418	ZINC01542529	ZINC01663914	ZINC02583275	ZINC05257628
ZINC00024610	ZINC00057423	ZINC01542530	ZINC01670696	ZINC02583278	ZINC05357597
ZINC00024912	ZINC00057424	ZINC01549072	ZINC01671135	ZINC02584638	ZINC05371120

Table 7.1.3 - virtual screening ligands included in the natural library. References (listed in full underneath the table) are given for the publication(s) linking the ligand and/or natural source to antitrypanosomal activity.

Name	Source	References
Lawsone (2-hydroxynapthoquinone)	*Lawsonia inermis*	Okpekon *et al.* (2004) (1)
Tingenone (Maitenin)	n.k.	Wright & Phillipson (1990) (2)
lucidamine A	*Garcinia lucida*	Mbaya & Ibrahim (2011) (3)
senecionine	*Crotalaria phillipsia*	Nibret et al (2009) (4)
furanolactone (columbin)	*Aristolchia albidia*	Nok *et al.* (2005) (5)
nimbin	*Azadirachta indica*	Mbaya *et al.* (2010) (6); Biswas *et al.* (2002) (7)
bruceine A	*Brucea javanica*	Bawm *et al.* (2009) (8)
bruceine B	*Brucea javanica*	Bawm *et al.* (2009) (8)
bruceine C	*Brucea javanica*	Bawm *et al.* (2009) (8)
brusatol	*Brucea javanica*	Bawm *et al.* (2009) (8)
bruceantinol	*Brucea javanica*	Bawm *et al.* (2009) (8)
bruceine D	*Brucea javanica*	Bawm *et al.* (2009) (8)
anthecotulide	*Anthemis auriculata*	Karioti *et al.* (2009) (8)
(-)-roemerine	*Annona senegalensis*	You *et al.* (1995) (9)
linalool	*Annona senegalensis*	Noudogbessi *et al.* (2011) (10)
(z)-b-ocimene	*Annona senegalensis*	Noudogbessi *et al.* (2011) (10)
(E)-b-ocimene	*Annona senegalensis*	Noudogbessi *et al.* (2011) (10)
Germacrene-D	*Annona senegalensis*	Noudogbessi *et al.* (2011) (10)
carryophyllene oxide	*Annona senegalensis*	Noudogbessi *et al.* (2011) (10)
beta carryophyllene	*Annona senegalensis*	Noudogbessi *et al.* (2011) (10)
palmitic acid	*Annona senegalensis*	Noudogbessi *et al.* (2011) (10)
caratuberside A	*Caralluma tuberculata*	Rizwani (1991) (11)
gallic acid	*Securidaca longepedunculata*	Muanda *et al.* (2010) (12)
chlorogenic acid	*Securidaca longepedunculata*	Muanda *et al.* (2010) (12)
caffeic acid	*Securidaca longepedunculata*	Muanda *et al.* (2010) (12)
epicatechin	*Securidaca longepedunculata*	Muanda *et al.* (2010) (12)
P-coumaric acid	*Securidaca longepedunculata*	Muanda *et al.* (2010) (12)
rutin	*Securidaca longepedunculata*	Muanda *et al.* (2010) (12)

Name	Source	References
Quercetin	*Securidaca longepedunculata*	Muanda *et al.* (2010) (12)
Quercetin 7-glucosyl	*Securidaca longepedunculata*	Muanda *et al.* (2010) (12)
Quercetin dehydrate	*Securidaca longepedunculata*	Muanda *et al.* (2010) (12)
cinnamic acid	*Securidaca longepedunculata*	Muanda *et al.* (2010) (12)
apigenin	*Securidaca longepedunculata*	Muanda *et al.* (2010) (12)
flazin	*Brucea javanica*	Su *et al.* (2002) (13)
yadanzioside A	*Brucea javanica*	Su *et al.* (2002) (13)
yadanzioside C	*Brucea javanica*	Su *et al.* (2002) (13)
azeleic acid	*Brucea javanica*	Su *et al.* (2002) (13)
(8S)-8-hydroxyhexadecanoic acid	*Brucea javanica*	Su *et al.* (2002) (13)
vanillin	*Brucea javanica*	Su *et al.* (2002) (13)
physcion	*Senna occidentalis*	Niranjan & Gupta (1973) (14)
emodin	*Senna occidentalis*	Niranjan & Gupta (1973) (14)
physcion 1-b-D-glucopyranoside	*Senna occidentalis*	Niranjan & Gupta (1973) (14)
B-sitosterol	*Senna occidentalis*	Niranjan & Gupta (1973) (14)
pinoresinol	*Brucea javanica*	Chen *et al.* (2009) (15)
cleomiscosin B	*Brucea javanica*	Chen *et al.* (2009) (15)
dihydrodehidrodiconiferyl alcohol	*Brucea javanica*	Chen *et al.* (2009) (15)
secoisolariciresinol	*Brucea javanica*	Chen *et al.* (2009) (15)
4-hydroxy-3-methoxybenzoic acid (vanillic acid)	*Brucea javanica*	Chen *et al.* (2009) (15)
3,4-dihydroxybenzoic acid	*Brucea javanica*	Chen *et al.* (2009) (15)
curcusone A	*Jatropha curcas*	Aiyelaagbe *et al.* (2011) (16)
curcusone B	*Jatropha curcas*	Aiyelaagbe *et al.* (2011) (16)
curcusone C	*Jatropha curcas*	Aiyelaagbe *et al.* (2011) (16)
curcusone D	*Jatropha curcas*	Aiyelaagbe *et al.* (2011) (16)
jatropholone	*Jatropha curcas*	Aiyelaagbe *et al.* (2011) (16)
Alpha-pinene	*Eucalyptus globulus*	http://www.essentialoils.co.za/essential-oils/eucalyptus.htm
Beta-pinene	*Eucalyptus globulus*	http://www.essentialoils.co.za/essential-oils/eucalyptus.htm
Alpha-phellandrene	*Eucalyptus globulus*	http://www.essentialoils.co.za/essential-oils/eucalyptus.htm
1,8-cineole	*Eucalyptus globulus*	http://www.essentialoils.co.za/essential-oils/eucalyptus.htm

Name	Source	References
limonene	*Eucalyptus globulus*	http://www.essentialoils.co.za/essential-oils/eucalyptus.htm
terpinen-4-ol	*Eucalyptus globulus*	http://www.essentialoils.co.za/essential-oils/eucalyptus.htm
aromadendrene	*Eucalyptus globulus*	http://www.essentialoils.co.za/essential-oils/eucalyptus.htm
epiglobulol	*Eucalyptus globulus*	http://www.essentialoils.co.za/essential-oils/eucalyptus.htm
globulol	*Eucalyptus globulus*	http://www.essentialoils.co.za/essential-oils/eucalyptus.htm
piperitone	*Eucalyptus globulus*	http://www.essentialoils.co.za/essential-oils/eucalyptus.htm
rhein	*Senna (Cassia) occidentalis*	Ibrahim *et al.* (2010) (17);Kaur & Ahmad (2014) (18)
aloe emodin	*Senna (Cassia) occidentalis*	Ibrahim *et al.* (2010) (17);Kaur & Ahmad (2014) (18)
chrysophanol	*Senna (Cassia) occidentalis*	Ibrahim *et al.* (2010) (17);Kaur & Ahmad (2014) (18)
chrysophanol-8-glucoside	*Senna (Cassia) occidentalis*	Ibrahim *et al.* (2010) (17);Kaur & Ahmad (2014) (18)
N-methyl morpholine	*Senna (Cassia) occidentalis*	Ibrahim *et al.* (2010) (17);Kaur & Ahmad (2014) (18)
daphneside	*Senna (Cassia) occidentalis*	Makhija *et al.* (2011)
daphnorin	*Senna (Cassia) occidentalis*	Makhija *et al.* (2011)
pinoresinol di-O-β-D-glucopyranoside	*Senna (Cassia) occidentalis*	Makhija *et al.* (2011)
lupeol	*Solanum schimperianum Hochst*	Al-Oqail *et al.* (2012) (19)
Beta-sitosterol glucoside	*Solanum schimperianum Hochst*	Al-Oqail *et al.* (2012) (19)
oleanolic acid	*Solanum schimperianum Hochst*	Al-Oqail *et al.* (2012) (19)
teferidine	*Solanum schimperianum Hochst*	Al-Oqail *et al.* (2012) (19)
teferin	*Solanum schimperianum Hochst*	Al-Oqail *et al.* (2012) (19)
ferutinin	*Solanum schimperianum Hochst*	Al-Oqail *et al.* (2012) (19)
retusin	*Solanum schimperianum Hochst*	Al-Oqail *et al.* (2012) (19)
kaempferol-3-O-glucopyranoside	*Solanum schimperianum Hochst*	Al-Oqail *et al.* (2012) (19)
acetamide	*Solanum schimperianum Hochst*	Al-Oqail *et al.* (2012) (19)
stearamide	*Solanum schimperianum Hochst*	Al-Oqail *et al.* (2012) (19)
Harman	*Guiera senegalensis*	Fiot *et al.* (2006) (20)
harmalan	*Guiera senegalensis*	Fiot *et al.* (2006) (20)
tetrahydroharman	*Guiera senegalensis*	Fiot *et al.* (2006) (20)
guieranone A	*Guiera senegalensis*	Fiot *et al.* (2006) (20)
cycloartenol	*Tamarindus indica L.*	Khanzada *et al.* (2008) (21)
Beta-amyrin	*Eucalyptus globulus*	Santos *et al.* (1997) (22)

Name	Source	References
erythrodiol	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
uvaol	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
acetyloleanolic acid	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
acetylbetulinic acid	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
acetylursolic acid	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
betulinic acid	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
skimiamine	*Zanthoxylum chalybeum*	Olila *et al.* (2002) (23)
nitogenin (diosgenin)	*Balanites aegyptiaca*	Dawidar & Fayez (1968) (24)
stigmasterol	*Eucalyptus globulus*	Yang & Guo (2007) (25)
euscaphic acid(1)	*Eucalyptus globulus*	Yang & Guo (2007) (25)
euscaphic acid(2)	*Eucalyptus globulus*	Yang & Guo (2007) (25)
2alpha-hydroxybetulinic acid	*Eucalyptus globulus*	Yang & Guo (2007) (25)
macrocarpal B	*Eucalyptus globulus*	Yang & Guo (2007) (25)
macrocarpal A	*Eucalyptus globulus*	Yang & Guo (2007) (25)
3,4,3'-tri-O-methylellagic acid	*Eucalyptus globulus*	Yang & Guo (2007) (25)
camaldulenside	*Eucalyptus globulus*	Yang & Guo (2007) (25)
ellagic acid	*Eucalyptus globulus*	Yang & Guo (2007) (25)
L-quebrachitol	*Paullinia pinnata*	Dongo *et al.* (2009) (26)
warburganal	*Warburgia salutaris (ugandensis)*	Mashimbye *et al.* (1999) (27)
polygodial	*Warburgia salutaris (ugandensis)*	Mashimbye *et al.* (1999) (27)
salutarisolide	*Warburgia salutaris (ugandensis)*	Mashimbye *et al.* (1999) (27)
muzigadial	*Warburgia salutaris (ugandensis)*	Mashimbye *et al.* (1999) (27)
ugandensidial	*Warburgia salutaris (ugandensis)*	Mashimbye *et al.* (1999) (27)
epipolygodial (isopolygodial)	*Warburgia salutaris (ugandensis)*	Mashimbye *et al.* (1999) (27)
mukaadial	*Warburgia salutaris (ugandensis)*	Mashimbye *et al.* (1999) (27)
zanhasaponin A	*Zanha africana*	Cuellar *et al.* (1997) (28)
zanhasaponin B	*Zanha africana*	Cuellar *et al.* (1997) (28)
pinitol	*Zanha africana*	Cuellar *et al.* (1997) (28)
filiformin	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
cathaformine	*Cassythia filiformis II*	Chang *et al.* (1998) (29)

Name	Source	References
cathafiline	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
actinodaphnine	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
predicentrine	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
ocoteine	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
(+)-syringaresinol	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
D-isoboldine	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
lysicamine	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
thalicmanine	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
stepharine	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
pronuciferine	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
O-methylflavinantine	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
( )-yangambin	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
isovanillin	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
Stigmasterol-D-glucoside	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
prosapogenin	*Foetidia africana*	Crublet *et al.* (2002) (30)
16-deoxybarringtogenol C	*Foetidia africana*	Crublet *et al.* (2002) (30)
Alpha-chaconine	*Solanum spp.*	Chataing *et al.* (1998) (31)
Alpha-solamargine (solasonine)	*Solanum spp.*	Chataing *et al.* (1998) (31)
tomatine	*Solanum spp.*	Chataing *et al.* (1998) (31)
7,8-epoxylignan	*Brucea javanica*	Chen *et al.* (2009) (15)
7'-hyroxy-lariciresionol	*Brucea javanica*	Chen *et al.* (2009) (15)
4-methoxyguaiacyglycerol	*Brucea javanica*	Chen *et al.* (2009) (15)
7-carbonylguaiacyglycerol	*Brucea javanica*	Chen *et al.* (2009) (15)
Brucojavan 1	*Brucea javanica*	Chen *et al.* (2009) (15)
Brucojavan 2	*Brucea javanica*	Chen *et al.* (2009) (15)
Brucojavan 3	*Brucea javanica*	Chen *et al.* (2009) (15)
isoplumbagin	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
lawsoniaside	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
lalioside	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
lawsoniaside B	*Lawsonia inermis*	Makhija *et al.* (2011) (32)

Name	Source	References
Agrimonolide 6-O-β-D-glucopyranoside	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
(+)-Syringaresinol O-β-D-glucopyranoside	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
Syringaresinol di-O-β-D-glucopyranoside	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
isoscutellarin	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
hennadiol	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
30-nor-lupan-3β-ol-20-one	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
botulin	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
laxanthone I	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
laxanthone II	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
laxanthone III	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
lacoumarin	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
apigenin-7-glucoside	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
luteolin	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
Alpha-terpineol	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
Alpha-ionone	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
Beta-ionone	*Lawsonia inermis*	Makhija *et al.* (2011) (32)
japodagrin	*Jatropha podagrica*	Aiyelaagbe *et al.* (2007) (33)
japodagrone	*Jatropha podagrica*	Aiyelaagbe *et al.* (2007) (33)
4Z-jatrogrossidentadion	*Jatropha podagrica*	Aiyelaagbe *et al.* (2007) (33)
15-epi-4Z-jatrogrossidentadion	*Jatropha podagrica*	Aiyelaagbe *et al.* (2007) (33)
2-hydroxyisojatrogrossidion	*Jatropha podagrica*	Aiyelaagbe *et al.* (2007) (33)
2-epihydroxyisojatrogrossidion	*Jatropha podagrica*	Aiyelaagbe *et al.* (2007) (33)
1-acetoxy-3,4,5-trimethoxyphenol	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
1-acetoxy-2,4,6-trimethoxyphenol	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
2,6-dimethoxy-p-benzoquinone	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
vomifoliol	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
Cis-p-methoxycinnamoyloxyursolicacid methyl ester	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
Trans-p-methoxycinnamoyloxyursolicacid methyl ester	*Eucalyptus globulus*	Santos *et al.* (1997) (22)

Name	Source	References
ursolic acid methyl ester	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
methyl 3beta,23-diacetoxy-12-ursen-28-oate	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
Cis-p-methoxycinnamoyloxyoleanolicacid methyl ester	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
Tetra-acetyldaucosterol	*Eucalyptus globulus*	Santos *et al.* (1997) (22)
3-O-Methylellagic acid 3'-O-alpha-4''-O-acetylrhamnopyranoside	*Eucalyptus globulus*	Yang & Guo (2007) (25)
etandrolide	*Entandrophragma angolense*	Orisadipe *et al.* (2005) (34)
3,23-Dioxotirucalla-7,24-dien-21-al	*Entandrophragma angolense*	Orisadipe *et al.* (2005) (34)
3,4-Secotirucalla-23-oxo-4(28),7,24-trien-21-al-3,21-dioic acid (21-methyl ester)	*Entandrophragma angolense*	Orisadipe *et al.* (2005) (34)
3,4-Secotirucalla-23-oxo-4(28),7,24-trien-21-al-3-oic acid	*Entandrophragma angolense*	Orisadipe *et al.* (2005) (34)
paullinamide A	*Paullinia pinnata*	Dongo *et al.* (2009) (26)
2-(4-hydroxy-3,5-dimethoxyphenyl)-3-hydroxymethyl-2,3-dihydro-1,4,5-trioxaphenanthren-6-one	*Paullinia pinnata*	Dongo *et al.* (2009) (26)
5alpha-poriferastane-3beta,6alpha-diol	*Paullinia pinnata*	Dongo *et al.* (2009) (26)
3alpha,7alpha-dideacetylkhivorin	*Khaya senegalensis*	Zhang *et al.* (2008) (35)
1alpha,3alpha,7alpha-trideacetylkhivorin	*Khaya senegalensis*	Zhang *et al.* (2008) (35)
mexicanoloid limonoid khayanone	*Khaya senegalensis*	Zhang *et al.* (2008) (35)
1-O-deacetylkhayanolide B	*Khaya senegalensis*	Zhang *et al.* (2008) (35)
khayanolide B	*Khaya senegalensis*	Zhang *et al.* (2008) (35)
khayanolide E	*Khaya senegalensis*	Zhang *et al.* (2008) (35)
1-O-deacetylkhayanolide E	*Khaya senegalensis*	Zhang *et al.* (2008) (35)
6-dehydroxykhayanolide E	*Khaya senegalensis*	Zhang *et al.* (2008) (35)
(-)-lyoniresinol	*Khaya senegalensis*	Zhang *et al.* (2008) (35)
(-)-lyoniresin-9-yl-b-D-xylopyranoside	*Khaya senegalensis*	Zhang *et al.* (2008) (35)
(-)-lyoniresin-4'-yl-b-D-glucopyranoside	*Khaya senegalensis*	Zhang *et al.* (2008) (35)

Name	Source	References
cassyformine	*Cassythia filiformis II*	Chang *et al.* (1998) (29)
D-mannitol	*Lawsonia inermis*	Chaudhary *et al.* (2010) (36)
triacontanol	*Lawsonia inermis*	Chaudhary *et al.* (2010) (36)
Apigenin-4'-glucoside	*Lawsonia inermis*	Chaudhary *et al.* (2010) (36)
luteolin-7-O-glucoside	*Lawsonia inermis*	Chaudhary *et al.* (2010) (36)
Luteolin-3'-O-glucoside	*Lawsonia inermis*	Chaudhary *et al.* (2010) (36)
cosmossin (acacetin-7-O-glucoside)	*Lawsonia inermis*	Chaudhary *et al.* (2010) (36)
acacetin	*Lawsonia inermis*	Chaudhary *et al.* (2010) (36)
fraxetin	*Lawsonia inermis*	Chaudhary *et al.* (2010) (36)
scopoletin	*Lawsonia inermis*	Chaudhary *et al.* (2010) (36)
esculetin	*Lawsonia inermis*	Chaudhary *et al.* (2010) (36)
1,2-dihydroxy- 4-o-glucosyloxy Naphthalene	*Lawsonia inermis*	Chaudhary *et al.* (2010) (36)
apiin	*Lawsonia inermis*	Chaudhary *et al.* (2010) (36)
3-methylnonacosane-1-ol	*Lawsonia inermis*	Chaudhary *et al.* (2010) (36)
stearic acid	*Lawsonia inermis*	Chaudhary *et al.* (2010) (36)
cinchonine	*Cinchona bark*	Hoet *et al.* (2004) (37);Merschjohann *et al.* (2001) (38)
quinine	*Cinchona bark*	Hoet *et al.* (2004) (37);Merschjohann *et al.* (2001) (38)
cinchonidine	*Cinchona bark*	Hoet *et al.* (2004) (37);Merschjohann *et al.* (2001) (38)
quinidine	*Cinchona bark*	Hoet *et al.* (2004) (37);Merschjohann *et al.* (2001) (38)
emetine	*Cephaelis ipecacuanha*	Hoet *et al.* (2004) (37);Merschjohann *et al.* (2001) (38)
berberine	*Many plant families*	Hoet *et al.* (2004) (37);Merschjohann *et al.* (2001) (38)
sanguinarine	*Many plant families*	Hoet *et al.* (2004) (37);Merschjohann *et al.* (2001) (38)
berbamine	*Many plant families, inc. Triclisia patens*	Hoet *et al.* (2004) (37);Merschjohann *et al.* (2001) (38)
thalisopidine	*Triclisia patens*	Hoet *et al.* (2004) (37);Camacho *et al.* (2002) (39)
pheanthine	*Triclisia patens*	Hoet *et al.* (2004) (37);Camacho *et al.* (2002) (39)
dioncophylline A	*Ancistrocladaceae, Dioncophyllaceae*	Hoet *et al.* (2004) (37);Bringmann et al (2003) (40)
dioncophylline B	*Ancistrocladaceae, Dioncophyllaceae*	Hoet *et al.* (2004) (37);Bringmann et al (2003) (40)
dioncophylline E	*Dioncophyllum thollonii*	Hoet *et al.* (2004) (37);Bringmann et al (2002) (41)
ancistroealaine A	*Ancistrocladus ealaensis*	Hoet *et al.* (2004) (37)
ancistroealaine B	*Ancistrocladus ealaensis*	Hoet *et al.* (2004) (37)

Name	Source	References
dioncopeltine A	*?*	Hoet *et al.* (2004) (37);Bringmann et al (2003) (40)
ancistrogriffithine A	*Ancistrocladus graffithii*	Hoet *et al.* (2004) (37);Bringmann et al (2002) (41)
O-methylmoschatoline	*Unonopsis buchtienii*	Hoet *et al.* (2004) (37);Waechter *et al.* (1999) (42)
lysicamine	*Unonopsis buchtienii*	Hoet *et al.* (2004) (37);Waechter *et al.* (1999) (42)
liriodenine	*Unonopsis buchtienii*	Hoet *et al.* (2004) (37);Waechter *et al.* (1999) (42)
cassythine	*Cassythia filiformis*	Hoet *et al.* (2004) (37)
dicentrine	*Cassythia filiformis*	Hoet *et al.* (2004) (37)
pancracine	*Narcissus angustifolius subsp. transcarpathicus*	Hoet *et al.* (2004) (37);Labraña *et al.* (2002) (43)
nangustine	*Narcissus angustifolius subsp. transcarpathicus*	Hoet *et al.* (2004) (37);Labraña *et al.* (2002) (43)
haemanthidine	*Cyrtanthus elatus*	Hoet *et al.* (2004) (37);Herrera *et al.* (2001) (44)
oxomaritidine	*Zephyranthes citrina*	Hoet *et al.* (2004) (37);Herrera *et al.* (2001) (45)
galanthine	*Zephyranthes citrina*	Hoet *et al.* (2004) (37);Herrera *et al.* (2001) (45)
lepadin D	n.k.	Hoet *et al.* (2004) (37)
lepadin E	n.k.	Hoet *et al.* (2004) (37)
lepadin F	n.k.	Hoet *et al.* (2004) (37)
fascaplysin	*Hyrtios erecta*	Hoet *et al.* (2004) (37)
ascididemin	*Marine metabolites*	Hoet *et al.* (2004) (37)
G2-bromoascicidemin	*Marine metabolites*	Hoet *et al.* (2004) (37)
cryptolepine	*Cryptolepis sanguinolenta*	Hoet *et al.* (2004) (37)
neocryptolepine	*Cryptolepis sanguinolenta*	Hoet *et al.* (2004) (37)
tryptanthrin	*Many plant species*	Hoet *et al.* (2004);Scovill *et al.* (2002)(46)
camptothecin	*Camptotheca acuminata*	Hoet *et al.* (2004) (37);Bodley & Shapiro (1995) (47)
ascofuranone	*Ascochyta visiae*	Hoet *et al.* (2004) (37)
ethyl gallate	*Many plant families*	Hoet *et al.* (2004) (37);Koide *et al.* (1998) (48)
punicalagin	*Combretum molle*	Hoet *et al.* (2004) (37);Asres *et al.* (2001) (49)
G7,8-dihydroxyflavone	n.k.	Hoet *et al.* (2004) (37)
quercetagetin	n.k.	Hoet *et al.* (2004) (37)
chrysosplenetin	*Ehretia amoena*	Hoet *et al.* (2004) (37)
chrysosplenol D	*Ehretia amoena*	Hoet *et al.* (2004) (37)

Name	Source	References
artemetin	*Ehretia amoena*	Hoet *et al.* (2004) (37)
demethylpraecansone B	*Tephrosia aequilata*	Hoet *et al.* (2004) (37)
praecansone B	*Tephrosia aequilata*	Hoet *et al.* (2004) (37)
cissampeloflavone	*Cissampelos pareira*	Hoet *et al.* (2004) (37)
amentoflavone	*Celaenodendron mexicanum*	Hoet *et al.* (2004) (31);Camacho *et al.* (2000) (50)
podocarpusflavone A	*Celaenodendron mexicanum*	Hoet *et al.* (2004) (37);Camacho *et al.* (2000) (50)
podocarpusflavone B	*Celaenodendron mexicanum*	Hoet *et al.* (2004) (37);Camacho *et al.* (2000) (50)
curcumin	*Curcuma longa*	Hoet *et al.* (2004) (37)
letestuianin C	*Aframomum letestuianum*	Hoet *et al.* (2004) (37)
G(4Z,6E)-5-hydroxy-1,7-bis(4-hydroxyphenyl)hepta-4,6-dien-3-one	*Aframomum letestuianum*	Hoet *et al.* (2004) (37)
letestuianin A	*Aframomum letestuianum*	Hoet *et al.* (2004) (37)
letestuianin B	*Aframomum letestuianum*	Hoet *et al.* (2004) (37)
justicidin B	*Phyllanthus piscatorum*	Hoet *et al.* (2004) (37)
piscatorin	*Phyllanthus piscatorum*	Hoet *et al.* (2004) (37)
plumbagin	*Drosera spp.*	Hoet *et al.* (2004) (37);Bringmann *et al.* (2003) (40)
Trans-isoshinanolone	*Drosera spp.*	Hoet *et al.* (2004) (37);Bringmann *et al.* (2003) (40)
diospyrin	*Diospyros montana*	Hoet *et al.* (2004) (37);Yardley *et al.* (1996) (51)
G2-(1-hydroxyethyl)-naphtho-[2,3-b]furan-4,9-quinone	*Kigelia pinnata*	Hoet *et al.* (2004) (37);Moideen *et al.* (1999) (52)
isopinnatal	*Kigelia pinnata*	Hoet *et al.* (2004) (37);Moideen *et al.* (1999) (52)
kigelinol	*Kigelia pinnata*	Hoet *et al.* (2004) (37);Moideen *et al.* (1999) (52)
isokigelinol	*Kigelia pinnata*	Hoet *et al.* (2004) (37);Moideen *et al.* (1999) (52)
knipholone	*Kniphofia foliosa*	Hoet *et al.* (2004) (37);Bringmann *et al.* (2002) (41)
knipholone anthrone	*Kniphofia foliosa*	Hoet *et al.* (2004) (37);Bringmann *et al.* (2002) (41)
G4'-O-demethylknipholone	*Kniphofia foliosa*	Hoet *et al.* (2004) (37);Bringmann *et al.* (2002) (41)
G4'-O-demethylknipholone-4'-O-B-d-glucopyranoside	*Bulbine frutescens*	Hoet *et al.* (2004) (37);Abegaz *et al.* (2002)
gaboroquinone A	*Bulbine frutescens*	Hoet *et al.* (2004);Abegaz *et al.* (2002)
gaboroquinone B	*Bulbine frutescens*	Hoet *et al.* (2004) (37);Abegaz *et al.* (2002)

Name	Source	References
azaanthroquinone	*Mitracarpus scaber*	Hoet *et al.* (2004) (37);Nok (2002) (53)
alloaromadendrine	*Essential oils*	Hoet *et al.* (2004) (37);Mikus *et al.* (2000) (54)
(-)-boscialin	*Boscia salicifolia*	Hoet *et al.* (2004) (37);Busch *et al.* (1998) (55)
ilimaquinone	*Dactylospongia elegans*	Hoet *et al.* (2004) (37)
pelorol	*Dactylospongia elegans*	Hoet *et al.* (2004) (37)
helenalin	*Arnica & Inula spp.*	Hoet *et al.* (2004) (37)
mexicanin I	*Arnica & Inula spp.*	Hoet *et al.* (2004) (37)
kolavenol	*Entada abyssinica*	Hoet *et al.* (2004) (37);Freiburghaus *et al.* (1997) (56)
G23-Hydroxy-5α-lanosta-7,9(11),24-triene-3-one	*Guarea rhophalocarpa*	Hoet *et al.* (2004) (37)
G5α-Lanosta-7,9(11),24-triene-3α,23-diol	*Guarea rhophalocarpa*	Hoet *et al.* (2004) (37)
G3-oxo-tirucalla-7,24Z-dien-26-oic acid	*Celaenodendron mexicanum*	Hoet *et al.* (2004) (37)
vernoguinosterol	*Vernonia guineensis*	Hoet *et al.* (2004) (37)
vernoguinoside	*Vernonia guineensis*	Hoet *et al.* (2004) (37)
klaivanolide	*Uvaria Klaineana*	Hoet *et al.* (2004) (37)
aculeatin D	*Amomum aculeatum*	Hoet *et al.* (2004) (37)
senegaline	*Annona senegalensis*	Hoet *et al.* (2004) (37)
squamocine	*Annona senegalensis*	Hoet *et al.* (2004) (37)
diallyl trisulfide	*Allium sativum*	Hoet *et al.* (2004) (37)
diallyl disulphide	*Allium sativum*	Hoet *et al.* (2004) (37);Nok *et al.* (1996) (57)
ajoene	*Allium sativum*	Hoet *et al.* (2004) (37)
manumycin A	*Streptomyces spp.*	Hoet *et al.* (2004) (37)
sinefungin	*Streptomyces grizeolus & S. incarnatus*	Hoet *et al.* (2004) (37)
O-methylancistrocladinine	*Ancistrocladus tarzanienses*	Gehrig & Efferth (2008) (58);Bringmann *et al.* (2004)
O,N-dimethylancistrocladine	*Ancistrocladus tarzanienses*	Gehrig & Efferth (2008) (58);Bringmann *et al.* (2004)
ancistrocladidine	*Ancistrocladus tarzanienses*	Bringmann *et al.* (2004)
ent-dioncophylleine A	*Ancistrocladus benomensis*	Gehrig & Efferth (2008) (58);Bringmann *et al.* (2005)
G5'-O-demethyl-ent-dioncophylleine A	*Ancistrocladus benomensis*	Gehrig & Efferth (2008);Bringmann *et al.* (2005)
dioncophylline D	*Ancistrocladus benomensis*	Gehrig & Efferth (2008);Bringmann *et al.* (2005)
G3-geranyloxy-6-methyl-1,8-dihydroxyanthraquinone	*Vismia orientalis*	Gehrig & Efferth (2008);Mbwambo *et al.* (2004)

Name	Source	References
bianthrone A1	*Vismia orientalis*	Gehrig & Efferth (2008);Mbwambo *et al.* (2004)
lanaroflavone	*Gingko biloba*	Gehrig & Efferth (2008);Wenigal *et al.* (2006)
bilobetin	*Gingko biloba*	Gehrig & Efferth (2008);Wenigal *et al.* (2006)
ginkgetin	*Gingko biloba*	Gehrig & Efferth (2008);Wenigal *et al.* (2006)
isoginkgetin	*Gingko biloba*	Gehrig & Efferth (2008);Wenigal *et al.* (2006)
sciadopitysin	*Gingko biloba*	Gehrig & Efferth (2008);Wenigal *et al.* (2006)
G3-hydroxyflavone	n.k.	Gehrig & Efferth (2008) (58); Tasdemir *et al.* (2006) (59)
rhamnetin	n.k.	Gehrig & Efferth (2008) (58); Tasdemir *et al.* (2006) (59)
G7,8,3',4'-tetrahydroxyflavone	n.k.	Gehrig & Efferth (2008) (58); Tasdemir *et al.* (2006) (59)
crypthophilic acid A	*Scrophularia cryptophila*	Gehrig & Efferth (2008) (58); Tasdemir *et al.* (2006) (59)
crypthophilic acid C	*Scrophularia cryptophila*	Gehrig & Efferth (2008) (58); Tasdemir *et al.* (2006) (59)
buddlejasaponin III	*Scrophularia cryptophila*	Gehrig & Efferth (2008) (58); Tasdemir *et al.* (2006) (59)
acetylastrogaloside I	*Astragalus baibutensis*	Gehrig & Efferth (2008)
astrogaloside I	*Astragalus baibutensis*	Gehrig & Efferth (2008)
astrogaloside II	*Astragalus baibutensis*	Gehrig & Efferth (2008)
astrogaloside III	*Astragalus baibutensis*	Gehrig & Efferth (2008)
harpagide	*Scrophularia cryptophila*	Gehrig & Efferth (2008) (58);Tasdemir *et al.* (2008) (60)
acetylharpagide	*Scrophularia cryptophila*	Gehrig & Efferth (2008) (58);Tasdemir *et al.* (2008) (60)
L-tryptophan	*Scrophularia cryptophila*	Gehrig & Efferth (2008) (58);Tasdemir *et al.* (2008) (60)
G7-deacetylgedunin	*Pseudocedrela kotschyi*	Gehrig & Efferth (2008)
G7-oxo-7-deacetylgedunin	*Pseudocedrela kotschyi*	Gehrig & Efferth (2008)
kotschyin A	*Pseudocedrela kotschyi*	Gehrig & Efferth (2008)
(E)-oroidin	*Agelas oroides*	Gehrig & Efferth (2008) (58);Tasdemir *et al.* (2007) (61)
G24-ethyl-cholest-5a-7-en-3-b-ol	*Agelas oroides*	Gehrig & Efferth (2008) (58);Tasdemir *et al.* (2007) (61)
G3-amino-1-(2-aminoimidazoyl)-prop-1-ene	*Agelas oroides*	Gehrig & Efferth (2008) (58);Tasdemir *et al.* (2007) (61)
G4,5-dibromopyrrole-2-carboxylic acid methyl ester	*Agelas oroides*	Gehrig & Efferth (2008) (58);Tasdemir *et al.* (2007) (61)
G4,5-dibromopyrrole-2-carboxylic acid	*Agelas oroides*	Gehrig & Efferth (2008) (58);Tasdemir *et al.* (2007) (61)
cynaropiricin	*Centaurea & Cynara spp.*	Adams *et al.* (2010) (62)
komaroviquinone	*Dracocephalum komarovi*	Suto *et al.* (2010) (63)
melampyroside	*Melampyrum arvense*	Kirmizibekmez *et al.* (2009) (64)

Name	Source	References
mussaenoside	*Melampyrum arvense*	Kirmizibekmez *et al.* (2009) (64)
mussaenocidic acid	*Melampyrum arvense*	Kirmizibekmez *et al.* (2009) (64)
G8-epi-loganin	*Melampyrum arvense*	Kirmizibekmez *et al.* (2009) (64)
benzoic acid	*Melampyrum arvense*	Kirmizibekmez *et al.* (2009) (64)
muqubilone B	*Diacamus bismarckensis*	Rubio *et al.* (2011) (65)
ent-muqubilone	*Diacamus bismarckensis*	Rubio *et al.* (2009) (65)
sigmosceptrellin A	*Diacamus bismarckensis*	Rubio *et al.* (2009) (65)
sigmosceptrellin A methyl ester	*Diacamus bismarckensis*	Rubio *et al.* (2009) (65)
sigmosceptrellin B	*Diacamus bismarckensis*	Rubio *et al.* (2009) (65)
epi-muqubillin A	*Diacamus bismarckensis*	Rubio *et al.* (2009) (65)
Epi-nuapapuin B methyl ester	*Diacamus bismarckensis*	Rubio *et al.* (2009) (65)
aerucyclamide B	*Microcystis aeruginosa*	Portman *et al.* (2011) (66)
aerucyclamide C	*Microcystis aeruginosa*	Portman *et al.* (2011) (66)
(-)-cubebin	*Zanthoxylum naranjillo*	Bastos *et al.* (1999) (67)
apicidin	*Fusarium pallidoroseum*	Murray *et al.* (2001) (68)
lunarine	*Lunaria biennis*	Hamilton *et al.* (2006) (69)
lunaridine	*Lunaria biennis*	Hamilton *et al.* (2006) (69)
tetrahydrolunarine	*Lunaria biennis*	Hamilton *et al.* (2006) (69)
lunarinol	*Lunaria biennis*	Hamilton *et al.* (2006) (69)
tetrahydrolunarinol	*Lunaria biennis*	Hamilton *et al.* (2006) (69)
numismine	*Lunaria biennis*	Hamilton *et al.* (2006) (69)
kukoamine A	*Lycium chinense & others*	Ponasik *et al.* (1995) (70)
coronaridine	*Muntafara sessilifolia*	Girardot *et al.* (2011) (71)
tabernaemontanine	*Muntafara sessilifolia*	Girardot *et al.* (2011) (71)
tabernaelegantine A	*Muntafara sessilifolia*	Girardot *et al.* (2012) (72)
tabernaelegantine B	*Muntafara sessilifolia*	Girardot *et al.* (2012) (72)
tabernaelegantine D	*Muntafara sessilifolia*	Girardot *et al.* (2012) (72)
violacein	*Chromobacterium violaceum*	Duran *et al.* (2011) (73)
miltirone	*Salvia miltiorrhiza*	Slusarczyk *et al.* (2011) (74)
methylenetanshinquinone	*Salvia miltiorrhiza*	Slusarczyk *et al.* (2011) (74)

Name	Source	References
hydrangdione A	*Salvia hydrangea*	Moridi Farimani *et al.* (2011) (75)
hydrangdione B	*Salvia hydrangea*	Moridi Farimani *et al.* (2011) (75)
arbusculin B	*Saussurea costus*	Julianti *et al.* (2011) (76)
a-cyclocostunolide	*Saussurea costus*	Julianti *et al.* (2011) (76)
costunolide	*Saussurea costus*	Julianti *et al.* (2011) (76)
dehydrocostuslactone	*Saussurea costus*	Julianti *et al.* (2011) (76)
parthenolide	n.k.	Julianti *et al.* (2011) (76)
zaluzanin D	*Laurus nobilis*	Julianti *et al.* (2011) (76)
eupatoriopicrin	*Eupatorium cannabinum*	Julianti *et al.* (2011) (76)
allaxanthone C	*Allanblackia monticola*	Lenta *et al.* (2007,a) (77);Azebaze *et al.* (2006) (78)
mangostin	*Allanblackia monticola*	Lenta *et al.* (2007,a) (77);Azebaze *et al.* (2006) (78)
tovophyllin A	*Allanblackia monticola*	Lenta *et al.* (2007,a) (77);Azebaze *et al.* (2006) (78)
rubraxanthone	*Allanblackia monticola*	Lenta *et al.* (2007,a) (77);Azebaze *et al.* (2006) (78)
norcowanin	*Allanblackia monticola*	Lenta *et al.* (2007,a) (77);Azebaze *et al.* (2006) (78)
friedelan-3-one	*Allanblackia monticola*	Lenta *et al.* (2007,a) (77);Azebaze *et al.* (2006) (78)
zenkequinone A	*Stereospermum zenkeri*	Lenta *et al.* (2007,a) (77);Lenta *et al.* (2007,b) (79)
zenkequinone B	*Stereospermum zenkeri*	Lenta *et al.* (2007,a) (77);Lenta *et al.* (2007,b) (79)
sterequinone F	*Stereospermum zenkeri*	Lenta *et al.* (2007,a) (77);Lenta *et al.* (2007,b) (79)
G6-O-methylcatalpol	n.k.	Ogungbe & Setzer (2009) (80)
G6-O-b-D-xylopyranosylaucubin	n.k.	Ogungbe & Setzer (2009) (80)
ajugol	n.k.	Ogungbe & Setzer (2009) (80)
ajugoside	n.k.	Ogungbe & Setzer (2009) (80)
aucubin	n.k.	Ogungbe & Setzer (2009) (80)
catalpol	n.k.	Ogungbe & Setzer (2009) (80)
ningpogenin	n.k.	Ogungbe & Setzer (2009) (80)
scrolepidoside	n.k.	Ogungbe & Setzer (2009) (80)
ancistrogriffine A	n.k.	Ogungbe & Setzer (2009) (80)
ancistrogriffine C	n.k.	Ogungbe & Setzer (2009) (80)
ancistrolikokine D	n.k.	Ogungbe & Setzer (2009) (80)
ancistrotanzanine A	n.k.	Ogungbe & Setzer (2009) (80)

Name	Source	References
ancistrotanzanine B	n.k.	Ogungbe & Setzer (2009) (80)
ancistrotanzanine C	n.k.	Ogungbe & Setzer (2009) (80)
ancistrotectonine	n.k.	Ogungbe & Setzer (2009) (80)
aromoline	n.k.	Ogungbe & Setzer (2009) (80)
korupensamine A	n.k.	Ogungbe & Setzer (2009) (80)
G3-O-acetylsanguinine	n.k.	Ogungbe & Setzer (2009) (80)
G8-Hydroxyheptadeca-1-ene-4,6-diyn-3-ylacetate	n.k.	Ogungbe & Setzer (2009) (80)
G8-Hydroxylheptadeca-4,6-diyn-3-ylacetate	n.k.	Ogungbe & Setzer (2009) (80)
G16-Acetoxy-11-hydroxyoctadeca-17-ene-12,14-diynyl acetate	n.k.	Ogungbe & Setzer (2009) (80)
G1,7-bis(4-hydroxyphenyl)-heptane-3,5-dione	n.k.	Ogungbe & Setzer (2009) (80)
G1,7-bis(4-hydroxyphenyl)-heptene-3,5-dione	n.k.	Ogungbe & Setzer (2009) (80)
ambigol A	n.k.	Ogungbe & Setzer (2009) (80)
ambigol C	n.k.	Ogungbe & Setzer (2009) (80)
angoroside C	n.k.	Ogungbe & Setzer (2009) (80)
chaetoxanthone A	n.k.	Ogungbe & Setzer (2009) (80)
chaetoxanthone B	n.k.	Ogungbe & Setzer (2009) (80)
letestuianin C	n.k.	Ogungbe & Setzer (2009) (80)
vismione D	n.k.	Ogungbe & Setzer (2009) (80)
G3-geranylemodin	n.k.	Ogungbe & Setzer (2009) (80)
arnicolide A	n.k.	Ogungbe & Setzer (2009) (80)
isoalantolactone	n.k.	Ogungbe & Setzer (2009) (80)
ivalin	n.k.	Ogungbe & Setzer (2009) (80)

References:

1. Okpekon T, Yolou S, Gleye C, Roblot F, Loiseau P, Bories C, et al. Antiparasitic activities of medicinal plants used in Ivory Coast. J Ethnopharmacol. 2004 Jan;90(1):91–7.
2. Wright CW, Phillipson JD. Natural products and the development of selective antiprotozoal drugs. Phytother Res. 1990 Aug 1;4(4):127–39.
3. Mbaya AW, Ibrahim UI. In vivo and in vitro Activities of Medicinal Plants on Haemic and Humoral Trypanosomes: A Review. Int J Pharmacol. 2011 Jan;7(1):1–11.
4. Nibret E, Sporer F, Asres K, Wink M. Antitrypanosomal and cytotoxic activities of pyrrolizidine alkaloid-producing plants of Ethiopia. J Pharm Pharmacol. 2009 Jun;61(6):801–8.
5. Nok AJ, Sallau BA, Onyike E, Useh NM. Columbin inhibits cholesterol uptake in bloodstream forms of *Trypanosoma brucei*-A possible trypanocidal mechanism. J Enzyme Inhib Med Chem. 2005 Jan 1;20(4):365–8.

6. Mbaya AW, Ibrahim UI, God OT, Ladi S. Toxicity and potential anti-trypanosomal activity of ethanolic extract of *Azadirachta indica* (Maliacea) stem bark: An in vivo and in vitro approach using *Trypanosoma brucei*. J Ethnopharmacol. 2010 Mar 24;128(2):495–500.
7. Biswas K, Chattopadhyay I, Banerjee RK, Bandyopadhyay U. Biological activities and medicinal properties of neem (*Azadirachta indica*). Curr Sci. 2002 Jun 10;82(11):1336–45.
8. Bawm S. Isolation of quassinoids from a medicinal plant, *Brucea javanica* and their in vitro activities against *Trypanosoma evansi*. J Vet Pharmacol Ther. 2009 Aug;32:129–265.
9. You M, Wickramaratne DBM, Silva GL, Chai H, Chagwedera TE, Farnsworth NR, et al. (-)-Roemerine, an Aporphine Alkaloid from *Annona senegalensis* That Reverses the Multidrug-Resistance Phenotype with Cultured Cells. J Nat Prod. 1995 Apr 1;58(4):598–604.
10. Noudogbessi J, Natta A, Avlessi F, Sohounhloué D, Figueredo G, Chalchat J. Chemical Composition of the Essential Oils Extracted from Two Annonaceae Required in Beninese Pharmacopeia. Aust J Basic Appl Sci. 2011 Feb;5(2):34.
11. Rizwani G. Phytochemical and studies on medicinal herbs, *Caralluma tuberculata* and *Caralluma edulis* [Internet]. University Of Karachi/ Department Of Pharmacognosy; 1991 [cited 2014 Sep 10]. Available from: http://eprints.hec.gov.pk/1241/1/958.html.htm
12. Muanda FN, Dicko A, Soulimani R. Assessment of polyphenolic compounds, in vitro antioxidant and anti-inflammation properties of *Securidaca longepedunculata* root barks. C R Biol. 2010;333(9):663–9.
13. Su B-N, Chang LC, Park EJ, Cuendet M, Santarsiero BD, Mesecar AD, et al. Bioactive constituents of the seeds of *Brucea javanica*. Planta Med. 2002 Aug;68(8):730–3.
14. Niranjan G, Gupta P. Chemical constituents of the flowers of *Cassia Occidentalis*. Planta Med. 2009 Jan;23(03):298–300.
15. Chen Q-J, Ouyang M-A, Tan Q-W, Zhang Z-K, Wu Z-J, Lin Q-Y. Constituents from the seeds of *Brucea javanica* with inhibitory activity of Tobacco mosaic virus. J Asian Nat Prod Res. 2009 Jun;11(6):539–47.
16. Aiyelaagbe OO, Hamid AA, Fattorusso E, Taglialatela-Scafati O, Schröder HC, Müller WEG. Cytotoxic Activity of Crude Extracts as well as of Pure Components from *Jatropha* Species, Plants Used Extensively in African Traditional Medicine. Evid Based Complement Alternat Med. 2011;
17. Ibrahim M, Aliyu A, Sallau A, Bashir M, Yunusa I, Umar T. *Senna occidentalis* leaf extract possesses antitrypanosomal activity and ameliorates the trypanosome-induced anemia and organ damage. Pharmacogn Res. 2010;2(3):175.
18. Kaur I, Ahmad S. Pharmacognosy, Phytochemistry and Pharmacology of *Cassia occidentalis* Linn. Int J Pharmacogn Phytochem Res. 2014;6(2):151–5.
19. Al-Oqail M, Hassan WHB, Ahmad MS, Al-Rehaily AJ. Phytochemical and biological studies of *Solanum schimperianum* Hochst. Saudi Pharm J. 2012 Oct;20(4):371–9.
20. Fiot J. Phytochemical and pharmacological study of roots and leaves of *Guiera senegalensis* JF Gmel (Combretaceae). J Ethnopharmacol. 2006;106(2):173.
21. Khanzada SK, Shaikh W, Sofia S, Kazi TG, Usmanghani K, Amina A, et al. Chemical constituents of *Tamarindus indica* L. medicinal plant in Sindh. Pak J Bot. 2008;40(6):2553–9.
22. Santos GG, Alves JCN, Rodilla JML, Duarte AP, Lithgow AM, Urones JG. Terpenoids and other constituents of *Eucalyptus globulus*. Phytochemistry. 1997 Apr;44(7):1309–12.
23. Olila D, Olwa-Odyek, Opuda-Asibo J. Screening of extracts of *Zanthoxylum chalybeum* and *Warburgia ugandensis* for activity against measles virus (Swartz and Edmonston strains) in vitro. Afr Health Sci. 2002 Apr;2(1):2–10.
24. Dawidar AAM, Fayez MBE. Steroid sapogenins--XIII: The constituents of B*alanites aegyptiaca*. Phytochemistry. 1969 Jan;8(1):261–5.
25. Yang X-W, Guo Q-M. Studies on chemical constituents in fruits of *Eucalyptus globulus*. Zhongguo Zhong Yao Za Zhi Zhongguo Zhongyao Zazhi China J Chin Mater Medica. 2007 Mar;32(6):496–500.
26. Dongo E, Hussain H, Miemanang RS, Tazoo D, Schulz B, Krohn K, et al. Chemical constituents of *Klainedoxa gabonenses* and *Paullinia pinnata*. Rec Nat Prod. 2009;3(3):165–9.
27. Mashimbye MJ, Maumela MC, Drewes SE. A drimane sesquiterpenoid lactone from *Warburgia salutaris*. Phytochemistry. 1999 Jun;51(3):435–8.
28. Cuéllar MJ, Giner RM, Recio MC, Just MJ, Máñez S, Cerdá M, et al. Zanhasaponins A and B, Antiphospholipase A2 Saponins from an Antiinflammatory Extract of *Zanha africana* Root Bark. J Nat Prod. 1997 Nov 1;60(11):1158–60.
29. Chang F-R, Chao Y-C, Teng C-M, Wu Y-C. Chemical Constituents from *Cassytha filiformis* II. J Nat Prod. 1998 Jul 1;61(7):863–6.
30. Crublet M-L, Pouny I, Delaude C, Lavaud C. Acylated triterpenoid saponins from the stem bark of *Foetidia africana*. J Nat Prod. 2002 Nov;65(11):1560–7.
31. Chataing B, Concepción JL, Lobatón R, Usubillaga A. Inhibition of Trypanosoma cruzi growth in vitro by *Solanum* alkaloids: a comparison with ketoconazole. Planta Med. 1998 Feb;64(1):31–6.
32. Makhija IK, Dhananjaya DR, Kumar VS, Devkar R, Khamar D, Manglani N, et al. *Lawsonia inermis* - From Traditional Use to Scientific Assessment. Afr J Pharm Sci Pharm [Internet]. 2011 Apr 25 [cited 2014 Sep 10];2(1). Available from: http://www.ajpspjournal.com/article/view/6816
33. Aiyelaagbe OO, Adesogan K, Ekundayo O, Gloer JB. Antibacterial diterpenoids from *Jatropha podagrica* Hook. Phytochemistry. 2007 Oct;68(19):2420–5.

34. Orisadipe AT, Adesomoju AA, D'Ambrosio M, Guerriero A, Okogun JI. Tirucallane triterpenes from the leaf extract of *Entandrophragma angolense*. Phytochemistry. 2005 Oct;66(19):2324–8.
35. Zhang H. Characterization of bioactive phytochemicals from the stem bark of African mahogany *Khaya senegalensis* (Meliaceae) [Internet]. Clemson University; 2008 [cited 2014 Sep 10]. Available from: http://gradworks.umi.com/33/44/3344280.html
36. Chaudhary G, Goyal S, Poonia P. *Lawsonia inermis* Linnaeus: A phytopharmacological review. Int J Pharm Sci Drug Res. 2010;2(2):91–8.
37. Hoet S, Opperdoes F, Brun R, Quetin-Leclercq J. Natural products active against African trypanosomes: a step towards new drugs. Nat Prod Rep. 2004 May 5;21(3):353–64.
38. Merschjohann K, Sporer F, Steverding D, Wink M. In Vitro Effect of Alkaloids on Bloodstream Forms of *Trypanosoma brucei* and *T. congolense*. Planta Med. 2001 Oct;67(7):623–7.
39. Camacho M del R, Phillipson JD, Croft SL, Rock P, Marshall SJ, Schiff PL. In vitro activity of *Triclisia patens* and some bisbenzylisoquinoline alkaloids against *Leishmania donovani* and*Trypanosoma brucei brucei*. Phytother Res. 2002 Aug;16(5):432–6.
40. Bringmann G, Hoerr V, Holzgrabe U, Stich A. Antitrypanosomal naphthylisoquinoline alkaloids and related compounds. Pharm. 2003 May;58(5):343–6.
41. Bringmann G, Messer K, Wolf K. ScienceDirect - Phytochemistry : Dioncophylline E from *Dioncophyllum thollonii*, the first 7,3′-coupled dioncophyllaceous naphthylisoquinoline alkaloid. Phytochemistry. 2002 Jul;60(4):389–97.
42. Waechter A-I, Cavé A, Hocquemiller R, Bories C, Muñoz V, Fournet A. Antiprotozoal activity of aporphine alkaloids isolated from *Unonopsis buchtienii* (Annonaceae). Phytother Res. 1999 Mar;13(2):175–7.
43. Labraña. ScienceDirect - Phytochemistry : Alkaloids from *Narcissus angustifolius* subsp. transcarpathicus (Amaryllidaceae). Phytochemistry. 2002 Aug;60(8):847–52.
44. Herrera. ScienceDirect - Fitoterapia : Alkaloids from *Cyrtanthus elatus*. Fitoterapia. 2001 May;72(4):444–8.
45. Herrera MR, Machocho AK, Brun R, Viladomat F, Codina C, Bastida J. Crinane and Lycorane Type Alkaloids from *Zephyranthes citrina*. Planta Med. 2001;67(2):191–3.
46. Scovill J, Blank E, Konnick M, Nenortas E, Shapiro T. Antitrypanosomal Activities of Tryptanthrins. Antimicrob Agents Chemother. 2002 Mar 1;46(3):882–3.
47. Bodley A, Shapiro T. Molecular and cytotoxic effects of camptothecin, a topoisomerase I inhibitor, on trypanosomes and Leishmania. Proc Natl Acad Sci U S A. 1995;92(9):3726–30.
48. Koide T, Nose M, Inoue M, Ogihara Y, Yabu Y, Ohta N. Trypanocidal Effects of Gallic Acid and Related Compounds. Planta Med. 2007 Jan;64(01):27–30.
49. Asres K, Bucar F, Knauder E, Yardley V, Kendrick H, Croft SL. In vitro antiprotozoal activity of extract and compounds from the stem bark of *Combretum molle*. Phytother Res. 2001 Nov;15(7):613–7.
50. Del Rayo Camacho M, Mata R, Castaneda P, Kirby GC, Warhurst DC, Croft SL, et al. Bioactive Compounds from *Celaenodendron mexicanum*. Planta Med. 2000 Jun;66(5):463–8.
51. Yardley V, Snowdon D, Croft S, Hazra B. In vitro Activity of Diospyrin and Derivatives against *Leishmania donovani, Trypanosoma cruzi* and*Trypanosoma brucei brucei.* Phytother Res. 1996 Nov;10(7):559–62.
52. Moideen SVK, Houghton PJ, Rock P, Croft SL, Aboagye-Nyame F. Activity of Extracts and Naphthoquinones from *Kigelia pinnata* Against *Trypanosoma brucei brucei* and *Trypanosoma brucei rhodesiense*. Planta Med. 1999 Aug;65(6):536–40.
53. Nok AJ. Azaanthraquinone inhibits respiration andin vitro growth of long slender bloodstream forms of *Trypanosoma congolense*. Cell Biochem Funct. 2002 Sep;20(3):205–12.
54. Mikus J, Harkenthal M, Steverding D, Reichling J. In vitro Effect of Essential Oils and Isolated Mono- and Sesquiterpenes on *Leishmania major* and *Trypanosoma brucei*. Planta Med. 2000 May;66(4):366–8.
55. Busch J, Grether Y, Ochs D, Séquin U. Total Synthesis and Biological Activities of (+)- and (−)-Boscialin and Their 1'-Epimers. J Nat Prod. 2011 Oct 25;61(5):591–7.
56. Freiburghaus F, Jonker S., Nkunya MH., Mwasumbi L., Brun R. In vitro trypanocidal activity of some rare Tanzanian medicinal plants. Acta Trop. 1997 Sep 30;67(3):181–5.
57. Nok AJ, Williams S, Onyenekwe PC. Allium sativum-induced death of African trypanosomes. Parasitol Res. 1996 Aug;82(7):634–7.
58. Gehrig S, Efferth T. Development of drug resistance in *Trypanosoma brucei rhodesiense* and *Trypanosoma brucei gambiense*. Treatment of human African trypanosomiasis with natural products (Review). Int J Mol Med. 2008 Oct;22(4):411–9.
59. Tasdemir D, Kaiser M, Brun R, Yardley V, Schmidt TJ, Tosun F, et al. Antitrypanosomal and Antileishmanial Activities of Flavonoids and Their Analogues: In Vitro, In Vivo, Structure-Activity Relationship, and Quantitative Structure-Activity Relationship Studies. Antimicrob Agents Chemother. 2006 Apr 1;50(4):1352–64.
60. Tasdemir D, Brun R, Franzblau SG, Sezgin Y, Calis I. Evaluation of antiprotozoal and antimycobacterial activities of the resin glycosides and the other metabolites of *Scrophularia cryptophila*. Phytomedicine Int J Phytother Phytopharm. 2008 Mar;15(3):209–15.

61. Tasdemir D, Topaloglu B, Perozzo R, Brun R, O'Neill R, Carballeira NM, et al. Marine natural products from the Turkish sponge Agelas oroides that inhibit the enoyl reductases from *Plasmodium falciparum, Mycobacterium tuberculosis* and *Escherichia coli*. Bioorg Med Chem. 2007 Nov 1;15(21):6834–45.
62. Adams M, Zimmermann S, Küenzi P, Julianti T, Hata Y, Brun R, et al. In vitro and in vivo activity of cynaropicrin against *Trypanosoma brucei rhodesiense*. Planta Med [Internet]. 2010 Aug [cited 2011 Nov 1];76(12). Available from: http://ovidsp.tx.ovid.com/sp-3.4.2a/ovidweb.cgi?&S=PJLHFPPAJJDDAFDLNCBLGELBHNHPAA00&Complete+Reference=S.sh.16%7c21%7c1
63. Suto Y, Kaneko K. A short and efficient asymmetric synthesis of komaroviquinone. Tetrahedron Lett - Tetrahedron Lett. 2010;51(48):6329–30.
64. Kırmızıbekmez H, Atay I, Kaiser M, Yeşilada E, Tasdemir D. Antiprotozoal activities of *Melampyrum arvense* and its secondary metabolites. Planta Med [Internet]. 2009 Jul [cited 2011 Nov 1];75(09). Available from: http://ovidsp.tx.ovid.com/sp-3.4.2a/ovidweb.cgi?&S=PJLHFPPAJJDDAFDLNCBLGELBHNHPAA00&Complete+Reference=S.sh.16%7c47%7c1
65. Rubio BK, Tenney K, Ang K-H, Abdulla M, Arkin M, McKerrow JH, et al. The Marine Sponge *Diacarnus bismarckensis* as a Source of Peroxiterpene Inhibitors of *Trypanosoma brucei*, the Causative Agent of Sleeping Sickness. J Nat Prod. 2011 Nov 1;72(2):218–22.
66. Portmann C, Blom JF, Kaiser M, Brun R, Jüttner F, Gademann K. Isolation of Aerucyclamides C and D and Structure Revision of Microcyclamide 7806A: Heterocyclic Ribosomal Peptides from *Microcystis aeruginosa* PCC 7806 and Their Antiparasite Evaluation. J Nat Prod. 2011 Nov 1;71(11):1891–6.
67. Bastos JK, Albuquerque S, Silva MLA. Evaluation of the Trypanocidal Activity of Lignans Isolated from the Leaves of *Zanthoxylum naranjillo*. Planta Med. 1999 Aug;65(6):541–4.
68. Murray PJ, Kranz M, Ladlow M, Taylor S, Berst F, Holmes AB, et al. The synthesis of cyclic tetrapeptoid analogues of the antiprotozoal natural product apicidin. Bioorg Med Chem Lett. 2001 Mar 26;11(6):773–6.
69. Hamilton CJ, Saravanamuthu A, Poupat C, Fairlamb AH, Eggleston IM. Time-dependent inhibitors of trypanothione reductase: analogues of the spermidine alkaloid lunarine and related natural products. Bioorg Med Chem. 2006 Apr 1;14(7):2266–78.
70. Ponasik JA, Strickland C, Faerman C, Savvides S, Karplus PA, Ganem B. Kukoamine A and other hydrophobic acylpolyamines: potent and selective inhibitors of *Crithidia fasciculata* trypanothione reductase. Biochem J. 1995 Oct 15;311(Pt 2):371–5.
71. Girardot M, Deregnaucourt C, Deville A, Dubost L, Joyeau R, Allorge L, et al. New vobasinyl-iboga bisindole alkaloids with antiparasitic activities from *Muntafara sessilifolia*. Planta Med [Internet]. 2011 Aug [cited 2011 Nov 2];77(12). Available from: http://ovidsp.tx.ovid.com/sp-3.4.2a/ovidweb.cgi?&S=GANFFPDMIMDDOEHONCBLEBFBJHPIAA00&Complete+Reference=S.sh.16%7c3%7c1
72. Girardot M, Deregnaucourt C, Deville A, Dubost L, Joyeau R, Allorge L, et al. Indole alkaloids from *Muntafara sessilifolia* with antiplasmodial and cytotoxic activities. Phytochemistry. 2012 Jan;73:65–73.
73. Durán M, Ponezi AN, Faljoni-Alario A, Teixeira MFS, Justo GZ, Durán N. Potential applications of violacein: a microbial pigment. Med Chem Res [Internet]. 2011 May [cited 2011 Nov 2]; Available from: http://www.springerlink.com/index/10.1007/s00044-011-9654-9
74. Ślusarczyk S, Zimmermann S, Kaiser M, Matkowski A, Hamburger M, Adams M. Antiplasmodial and Antitrypanosomal Activity of Tanshinone-Type Diterpenoids from *Salvia miltiorrhiza*. Planta Med. 2011 Mar;77(14):1594–6.
75. Moridi Farimani M, Bahadori B, Taheri S, Nejad Ebrahimi S, Hamburger M. Antiplasmodial and antitrypanosomal triterpenoids from *Salvia hydrangea* with rare carbon skeletons. Planta Med [Internet]. 2011 Aug [cited 2011 Nov 3];77(12). Available from: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0031-1282576
76. Julianti T, Hata Y, Zimmermann S, Kaiser M, Hamburger M, Adams M. Antitrypanosomal sesquiterpene lactones from *Saussurea costus*. Fitoterapia. 2011 Oct;82(7):955–9.
77. Lenta BN, Vonthron-Sénécheau C, Weniger B, Devkota KP, Ngoupayo J, Kaiser M, et al. Leishmanicidal and cholinesterase inhibiting activities of phenolic compounds from *Allanblackia monticola* and *Symphonia globulifera*. Mol Basel Switz. 2007;12(8):1548–57.
78. Azebaze AGB, Meyer M, Valentin A, Nguemfo EL, Fomum ZT, Nkengfack AE. Prenylated xanthone derivatives with antiplasmodial activity from *Allanblackia monticola* STANER L.C. Chem Pharm Bull (Tokyo). 2006 Jan;54(1):111–3.
79. Lenta BN, Weniger B, Antheaume C, Noungoue DT, Ngouela S, Assob JCN, et al. Anthraquinones from the stem bark of *Stereospermum zenkeri* with antimicrobial activity. Phytochemistry. 2007 Jun;68(11):1595–9.
80. Ogungbe IV, Setzer WN. Comparative Molecular Docking of Antitrypanosomal Natural Products into Multiple *Trypanosoma brucei* Drug Targets. Molecules. 2009 Apr 14;14:1513–36.

Table 7.1.4 - substances used for the treatment of trypanosomiasis(1). The compounds were included in the 'Knowledge' library for virtual screening. Examples are historically used, currently used, or may be used in the future.

Nagana Red
Trypan Red
Trypan Blue
Afridol Violet
Suramin
Atoxyl
arsenophenylglycine
Tryparsamide
Melarsen
Melarsen Oxide
Melarsoprol
Synthalin
Diamidino-1,11-n-undecane
Stilbamidine
Pentamidine
2,5-bis(4-aminophenyl)-furan
Nifurtimox
Eflornithine
Disulfuram
N-ethylmaleimide
Benznidazole

Reference:

1. Steverding D. The development of drugs for treatment of sleeping sickness: a historical review. Parasit Vectors [Internet]. 2011 [cited 2011 Sep 27];3(15). Available from: http://www.parasitesandvectors.com/content/3/1/15

Table 7.1.5 - proteins identified by protein structure similarity clustering. The Q-score is a measure of structural similarity that takes into account the number of residues aligned. The maximum value for a Q-score is (resulting from two identical proteins). The P-score represents the probability of finding a better alignment by chance. All results with a P-score lower than P were rejected. Only results with a percentage sequence similarity (%ss) less than 20 were accepted.

Protein	ID	Q-score	P-score	RMSD	%ss	Relevance
Human GalE	d1hzjb_ 1HZJ:B	0.272	8.35	1.919	12.8	Galactose metabolism
E. coli GalE	d1naia_	0.264	7.77	1.988	12.9	Galactose metabolism
Spinach SQD1 protein complex	d1qrra_	0.207	7.11	2.316	12.9	Phosphate stress in photosynthetic membranes
Strep. Venezuelae dTDP-glucose 4,6-dehydratase	d1r6da_	0.266	2.03	2.213	14.8	6-deoxyhexose synthesis (cell wall glycans)
Arabidopsis thaliana GDP-mannose 4,6-dehydratase	d1n7hb_	0.260	6.24	2.245	11.9	Mannose & fructose metabolism
Salmonella CDP-tyvelose 2-epimerase	d1orrb_	0.263	4.37	1.827	15.4	tyvelose (O-antigens of some bacteria) synthesis
Human GDP-mannose 4,6-dehydratase	d1t2daa_	0.277	4.41	2.181	14.8	Mannose & fructose metabolism
E. coli ArnA decarboxylase domain	d1u9ja_	0.230	4.22	2.438	13.0	Modification of Lipid A of LPS – cause of polymixin and cationic antimicrobial peptide (CAMP) resistance.
P. aeruginosa GDP-4-keto-6-deoxy-D-mannose reductase	2pk3:B	0.289	6.55	2.122	8.7	L-fucose pathway. L-fucose is in bacterial cell wall polysaccharide, LPS, human selectins – role in inflammation and metastasis.
Bacillus Cereus UDP-N-acetyl-glucosamine-4-epimerase	3m2p:E	0.225	4.16	2.022	17.5	Initiation of O-glycosylation
E. coli K-12 ADP-glycero-D-mannoheptose 6-epimerase	1eq2:F	0.225	3.92	2.364	11.7	ADP-D-glycero-D-mannoheptose and ADP-L-glycero-D-mannoheptose interconversion – LPS synthesis
Human FX protein (GDP-L-fucose synthase)	4e5y:A	0.241	3.62	2.329	9.2	Human antigens and selectins – role in inflammation and metastasis; leukocyte adhesion.
Salmonella typhii CDP-D-glucose 4,6-dehydratase	1wvg:A	0.525	21.58	1.913	17.4	First step in tyvelose synthesis – O antigens
Trypanosoma brucei GalE	1gy8:A	0.483	17.97	2.071	17.6	Galactose metabolism, validated target
Arabidopsis thaliana GDP-mannose 3',5'-epimerase	2c59:B	0.471	10.03	2.037	16.7	Possible precursors for plant vitamin C biosynthesis.
Salmonella enterica typhimurium dTDP-6-deoxy-L-lyxo-4-hexulose reductase (dTDP-4-	1n2s:A	0.430	7.10	2.589	15.7	L-rhamnose synthetic pathway – bacterial cell-wall anchoring and LPS.

Protein	ID	Q-score	P-score	RMSD	%ss	Relevance
dehydrorhamnose reductase)						
Grape (*Vitis vinifera*) dihydroflavonol 4-reductase (dihydrokaempferol reductase)	2c29:D	0.421	15.05	2.396	16.1	Flavonoid (anthocyanin and condensed tannin) synthesis – plant survival.
Staph Aureus capsular polysaccharide synthesizing enzyme (CAPE)	3w1v:B	0.335	12.09	2.264	15.8	Obvious drug target
Digitalis lanata progesterone 5β-reductase	2v6f:A	0.323	9.11	2.738	10.4	digitalis cardenolide synthesis
Human S-adenosyl-methionine synthetase, subunit beta	2ydx:A	0.452	11.85	2.248	18.1	Essential to normal cell function and growth. Cancer drug target. Note two types: liver- & non-liver- (kidney-) specific.
Human UDP-xylose synthase (UDP-glucoronic acid decarboxylase)	4gll:B	0.452	14.14	2.117	19.8	Polyamine & proteoglycan synthesis – cell/tumor growth.

Table 7.1.6 - proteins and ligands identified using PSSC. The ligands listed were included in the virtual screen against TDH. The references are listed below the table.

Protein	Organism	Ligands
UDP-galactose 4'-epimerase (GalE)	*Saccharomyces fragilis*	phenylglyoxal; 1,2-cyclohexanediol; 2,3-butanedione, 5'UMP(1); DTNB(2).
	Kluyveromyces marxianus	2',3'-O-(2,4,6-Trinitrocyclohexadienylidene)uridine 5'-monophosphate (3); diethyldicarbonate (4); L-arabinose(5), UTP(6).
	Bos taurus	2-Hydroxy-5-nitrobenzyl bromide; showdomycin(7).
	Saccharomyces cerevisiae	D-xylose, L-arabinose(8).
dTDP-glucose 4,6-dehydratase	*E. coli*	dTDP-6-deoxy-D-glucose; dTDP-6-deoxy-D-galactose(9). dTDP-xylose(10).
	Streptomyces sp. C5	TMP, TDP, TTP(11).
GDP-glucose 4,6-dehydratase	*Phaseolus vulgaris*	GDP-D-glucose, GMP, GTP, ITP(12).
UDP-N-acetyl-glucosamine-4-epimerase	*Homo sapiens*	6-Bromo-2-hydroxy-naphthalene-1-carbaldehyde O-[5-(2,4-dioxo-3,4-dihydro-2H-pyrimidin-1-yl)-3,4-dihydroxy-tetrahydro-furan-2-ylmethyl]-oxime(13).
GDP-L-fucose synthetase	*E. coli*	GDP, GDP-fucose(14).
CDP-D-glucose 4,6-dehydratase	*Yersinia pseudotuberculosis*	CDP-6-deoxy-6,6-difluoro-d-glucose(15).
Dihydroflavonol 4-reductase	*Onobrychis viciifolia*	DTT, diethyldithiocarbamate(16).
	Salmonella enterica	dTDP-6-deoxy-L-mannose(17),
	Mycobacterium tuberculosis	CHEMBL375328(NSC 9037), cid 72722(Chembridge 5108509), CHEMBL175266(NSC 5426), CHEMBL205119(NSC 371300) (18).
UDP-D-xylose synthase (UDP-glucuronic acid decarboxylase)	*Cryptococcus laurentii*	3-acetylpyridine, 3-formylpyridine, 3-propionylpyridine, AMP, UDP, UDP-glucose, UDP-xylose, UMP(19)
	Oryza sativa	CDP, CTP, GDP, GDP-mannose, TDP, TMP, TTP, UDP-glucose, UDP-xylose(20).
	Hordeum vulgare	UDP-galactose, UDP-glucose, UDP-D-xylose(21).
	Triticum aestivum	UDP, UDP-arabinose, UMP(22).
S-adenosylmethionine synthetase	*Rattus norvegicus*	(2S,4S)-amino-4,5-epoxypentanoic acid (AEP)(23). L-buthionine-(S,R)-sulfoximine; S-nitrosoglutathione; S-nitrosoglutathione monoethyl

		ester(24). Epoxy analogue I(a); epoxy analogue I(b); epithioaminoacid analogue II(a); epoxy analogue I(c); epithioaminoacid analogue II(b); epithioaminoacid analogue II(c)(25). CHEMBL608501(26). CHEMBL609500(27). CHEMBL603986(28). CHEMBL608502(26). CHEMBL1791426(27). CHEMBL609078(29). CHEMBL606493(28). CHEMBL611540(27). CHEMBL1791425(27). AMPPNP(30). CHEMBL1791411(28). CHEMBL1791415(31). CHEMBL607665(29). CHEMBL1791416(29).
	E. coli	1-(3-(2-ethoxyphenyl)ureidoacetyl)-4-(2-methyl-5-nitrophenyl)semicarbazide; 1-(4-chloro-2-nitrophenyl)-3-(4-sulfamoylphenyl)-urea(32). Adenyl-5'-ylimidodiphosphate; GTP; S-carbamylcysteine(33). Diimidotriphosphate(34).
	Mycobacterium smegmatis	1-methyluric acid; 1-methylxanthine; 2,6-diaminopurine; 2,6-dichloropurine; 2-amino-6-carboxyethylmercaptopurine; 2-amino-6-chloropurine riboside; 2-amino-6-chloropurine-9-acetic acid; 2-aminopurine; 2-hydroxypurine; 3,7-dimethyluric acid; 6-benzyloxypurine; 6-bromopurine; 6-chloropurine; 6-chloropurine riboside; 6-cyanopurine; 6-dimethylallylaminopurine riboside; 6-dimethylaminopurine; 6-mercaptopurine; 6-mercaptopurine riboside; 6-propoxypurine; 7-hydroxypropyl theophylline; 7-methyluric acid; 7-methylxanthine; 8-aza-2,6-diaminopurine; 8-azaguanine; 8-chlorotheophylline; alpha-methyl-DL-methionine; azathioprine; L-ethioninamide; L-ethionine; L-methionine methyl ester; L-methionine sulfone; L-methionine sulfoxide; L-methionine sulfoxime; L-penicillamine; methylthio propionaldehyde; O-methylguanine; uric acid; xanthine (35).

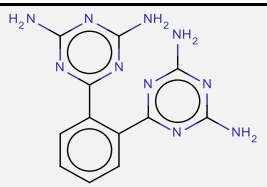
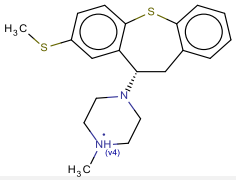
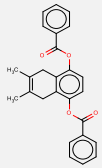
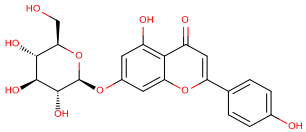
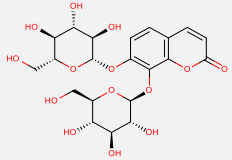
	Catharanthus roseus	ADP; CTP; ethionine; L-ornithine; GTP; putrescine; spermidine; spermine(36).
	Saccharomyces cerevisiae	Alpha,beta-methylene-ATP; beta,gamma-methylene-ATP; CTP; dGTP; GDP; GTP; TTP; UTP(37). AMP(38). 1-aminocyclopentanecarboxylic acid; DL-2-Amino-trans-4-hexenoic acid; GTP; S-Trifluoromethyl-L-homocysteine(39).
	Trypanosoma brucei brucei	Cycloleucine; D-methionine; DL-ethionine; L-2-Amino-4-methoxy-cis-but-3-enoic acid; S-adenosyl-L-homocysteine; S-adenosylmethionine; seleno-L-methionine(40).
	Mus musculus	L-2-Amino-4-methoxy-cis-but-3-enoic acid; L-2-Amino-4-methylthio-cis-but-3-enoic acid(41).
	Arabidopsis thaliana	nitrosoglutathione(42).
	Leishmania infantum	S-adenosyl L-ethionine(43).

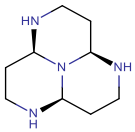
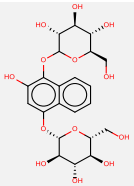
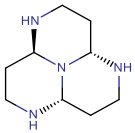
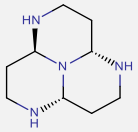
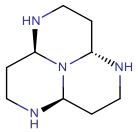
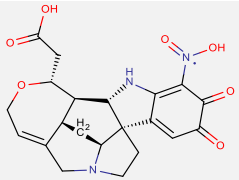
References:
1. Mukherji S, Bhaduri A. UDP-glucose 4-epimerase from *Saccharomyces fragilis*. Presence of an essential arginine residue at the substrate-binding site of the enzyme. J Biol Chem. 1986 Apr 5;261(10):4519–24.
2. Ray M, Kar K, Bhaduri A. UDP-glucose 4-epimerase from *Saccharomyces fragilis*. Involvement of sulfhydryl group(s) at the active site. Biochim Biophys Acta. 1978 Oct 12;526(2):635–9.
3. Ray S, Ali E, Bhaduri A. Microenvironment at the substrate binding subsite of the active site of UDPglucose 4-epimerase from *Kluyveromyces fragilis* using a fluorescent analog of UMP. Indian J Biochem Biophys. 1992 Apr;29(2):209–13.
4. Mukherji S, Bhaduri A. An essential histidine residue for the activity of UDPglucose 4-epimerase from *Kluyveromyces fragilis*. J Biol Chem. 1992 Jun 15;267(17):11709–13.
5. Majumdar S, Bhattacharjee H, Bhattacharyya D, Bhaduri A. UDP-galactose 4-epimerase from *Kluyveromyces fragilis*: reconstitution of holoenzyme structure after dissociation with parachloromercuribenzoate. Eur J Biochem FEBS. 1998 Oct 15;257(2):427–33.
6. Brahma A, Mandal C, Bhattacharyya D. Characterization of a dimeric unfolding intermediate of bovine serum albumin under mildly acidic condition. Biochim Biophys Acta BBA - Proteins Proteomics. 2005 Aug 10;1751(2):159–69.
7. Geren CR, Geren LM, Ebner KE. Inhibition and inactivation of bovine mammary and liver UDP-galactose-4-epimerases. J Biol Chem. 1977 Mar 25;252(6):2089–94.
8. Cármenes RS, Gascón S, Moreno F. Mechanism of inactivation of UDP-glucose 4-epimerase from *Saccharomyces cerevisiae* by D-xylose and L-arabinose. Yeast Chichester Engl. 1986 Jun;2(2):101–8.

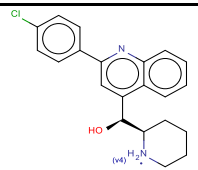
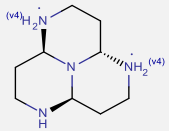
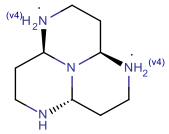
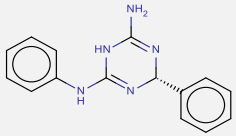
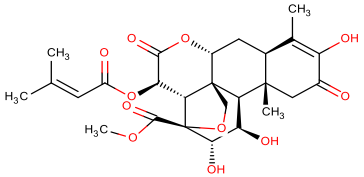
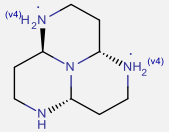
9. Zarkowsky H, Lipkin E, Glaser L. The mechanism of 6-deoxyhexose synthesis IV. The role of pyridine nucleotide in substrate release. Biochem Biophys Res Commun. 1970 Feb 20;38(4):787–93.
10. Hegeman AD, Gross JW, Frey PA. Probing catalysis by *Escherichia coli* dTDP-glucose-4,6-dehydratase: identification and preliminary characterization of functional amino acid residues at the active site. Biochemistry (Mosc). 2001 Jun 5;40(22):6598–610.
11. Thompson MW, Strohl WR, Floss HG. Purification and characterization of TDP-D-glucose 4,6-dehydratase from anthracycline-producing streptomycetes. J Gen Microbiol. 1992 Apr;138(4):779–86.
12. Liao TH, Barber GA. Purification of guanosine 5'-diphosphate D-mannose oxidoreductase from *Phaseolus vulgaris*. Biochim Biophys Acta. 1972 Jul 13;276(1):85–93.
13. Winans KA, Bertozzi CR. An Inhibitor of the Human UDP-GlcNAc 4-Epimerase Identified from a Uridine-Based Library: A Strategy to Inhibit O-Linked Glycosylation. Chem Biol. 2002 Jan;9(1):113–29.
14. Menon S, Stahl M, Kumar R, Xu GY, Sullivan F. Stereochemical course and steady state mechanism of the reaction catalyzed by the GDP-fucose synthetase from Escherichia coli. J Biol Chem. 1999 Sep 17;274(38):26743–50.
15. Chang C-WT, Chen XH, Liu H. CDP-6-deoxy-6,6-difluoro-d-glucose: A Mechanism-Based Inhibitor for CDP-d-glucose 4,6-Dehydratase. J Am Chem Soc. 1998 Sep 1;120(37):9698–9.
16. Singh S, McCallum J, Gruber MY, Neil Towers GH, Muir AD, Bohm BA, et al. Biosynthesis of flavan-3-ols by leaf extracts of *Onobrychis viciifolia*. Phytochemistry. 1997 Feb;44(3):425–32.
17. Elling L, Rupprath C, Günther N, Römer U, Verseck S, Weingarten P, et al. An enzyme module system for the synthesis of dTDP-activated deoxysugars from dTMP and sucrose. Chembiochem Eur J Chem Biol. 2005 Aug;6(8):1423–30.
18. Wang Y, Hess TN, Jones V, Zhou JZ, McNeil MR, Andrew McCammon J. Novel inhibitors of *Mycobacterium tuberculosis* dTDP-6-deoxy-l-lyxo-4-hexulose reductase (RmlD) identified by virtual screening. Bioorg Med Chem Lett. 2011 Dec 1;21(23):7064–7.
19. Ankel H, Feingold DS. [50] UDP-glucuronic acid decarboxylase. In: Elizabeth F. Neufeld VG, editor. Methods in Enzymology [Internet]. Academic Press; 1966 [cited 2013 Jul 10]. p. 287–92. Available from: http://www.sciencedirect.com/science/article/pii/0076687966080546
20. Suzuki K, Watanabe K, Masumura T, Kitamura S. Characterization of soluble and putative membrane-bound UDP-glucuronic acid decarboxylase (OsUXS) isoforms in rice. Arch Biochem Biophys. 2004 Nov 15;431(2):169–77.
21. Zhang Q, Shirley N, Lahnstein J, Fincher GB. Characterization and expression patterns of UDP-D-glucuronate decarboxylase genes in barley. Plant Physiol. 2005 May;138(1):131–41.
22. Ankel H, Feingold DS. Biosynthesis of Uridine Diphosphate D-Xylose. I. Uridine Diphosphate Glucuronate Carboxy-lyase of Wheat Germ*. Biochemistry (Mosc). 1965 Nov 1;4(11):2468–75.
23. González B, Pajares MA, Hermoso JA, Guillerm D, Guillerm G, Sanz-Aparicio J. Crystal structures of methionine adenosyltransferase complexed with substrates and products reveal the methionine-ATP recognition and give insights into the catalytic mechanism. J Mol Biol. 2003 Aug 8;331(2):407–16.
24. Corrales FJ, Pérez-Mato I, Sánchez Del Pino MM, Ruiz F, Castro C, García-Trevijano ER, et al. Regulation of mammalian liver methionine adenosyltransferase. J Nutr. 2002 Aug;132(8 Suppl):2377S – 2381S.
25. Lavrador K, Allart B, Guillerm D, Guillerm G. A New Series of S-Adenosyl-L-Methionine Synthetase Inhibitors. J Enzyme Inhib Med Chem. 1998 Jan;13(5):361–7.
26. Vrudhula VM, Kappler F, Afshar C, Ginell SL, Lessinger L, Hampton A. Approaches to isozyme-specific inhibitors. 16. A novel methyl-C5' covalent adduct of L-ethionine and beta,gamma-imido-ATP as a potent multisubstrate inhibitor of rat methionine adenosyltransferases. J Med Chem. 1989 Apr;32(4):885–90.

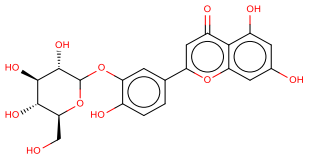
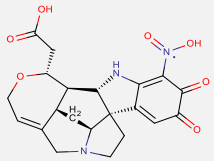
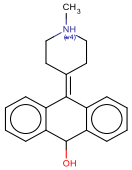
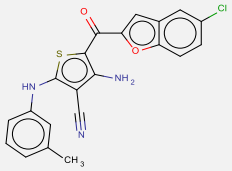
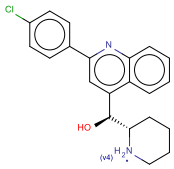
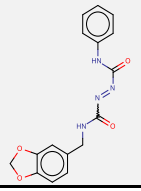
27. Kappler F, Vrudhula VM, Hampton A. Toward the synthesis of isozyme-specific enzyme inhibitors. Potent inhibitors of rat methionine adenosyltransferases. Effect of one-atom elongation of the ribose-P alpha bridge in two covalent adducts of L-methionine and beta,gamma-imido-ATP. J Med Chem. 1988 Feb;31(2):384–9.
28. Kappler F, Vrudhula VM, Hampton A. Isozyme-specific enzyme inhibitors. 14. 5'(R)-C-[(L-homocystein-S-yl)methyl]adenosine 5'-(beta,gamma-imidotriphosphate), a potent inhibitor of rat methionine adenosyltransferases. J Med Chem. 1987 Sep;30(9):1599–603.
29. Vrudhula VM, Kappler F, Hampton A. Isozyme-specific enzyme inhibitors. 13. S-[5'(R)-[(N-triphosphoamino)methyl]adenosyl]-L-homocysteine, a potent inhibitor of rat methionine adenosyltransferases. J Med Chem. 1987 May;30(5):888–94.
30. Kappler F, Hai TT, Hampton A. Isozyme-specific enzyme inhibitors. 10. Adenosine 5'-triphosphate derivatives as substrates or inhibitors of methionine adenosyltransferases of rat normal and hepatoma tissues. J Med Chem. 1986 Mar;29(3):318–22.
31. Kappler F, Hai TT, Cotter RJ, Hyver KJ, Hampton A. Isozyme-specific enzyme inhibitors. 11. L-homocysteine-ATP S-C5' covalent adducts as inhibitors of rat methionine adenosyltransferases. J Med Chem. 1986 Jun;29(6):1030–8.
32. Taylor JC, Bock CW, Takusagawa F, Markham GD. Discovery of novel types of inhibitors of S-adenosylmethionine synthesis by virtual screening. J Med Chem. 2009 Oct 8;52(19):5967–73.
33. Markham GD, Hafner EW, Tabor CW, Tabor H. S-Adenosylmethionine synthetase from *Escherichia coli.* J Biol Chem. 1980 Oct 10;255(19):9082–92.
34. Markham GD, Reczkowski RS. Structural studies of inhibition of S-adenosylmethionine synthetase by slow, tight-binding intermediate and product analogues. Biochemistry (Mosc). 2004 Mar 30;43(12):3415–25.
35. Berger BJ, Knodel MH. Characterisation of methionine adenosyltransferase from *Mycobacterium smegmatis* and *M. tuberculosis*. BMC Microbiol. 2003 Jun 16;3:12.
36. Schröder G, Eichel J, Breinig S, Schröder J. Three differentially expressed S-adenosylmethionine synthetases from *Catharanthus roseus*: molecular and functional characterization. Plant Mol Biol. 1997 Jan;33(2):211–22.
37. Cho TC, Talalay P. Inhibition of ATP: L-methionine S-adenosyltransferase of bakers' yeast by structural analogues of ATP. Biochim Biophys Acta. 1973 Oct 10;321(2):467–74.
38. Luo Y, Yuan Z, Luo G, Zhao F. Expression of secreted His-tagged S-adenosylmethionine synthetase in the methylotrophic yeast *Pichia pastoris* and its characterization, one-step purification, and immobilization. Biotechnol Prog. 2008 Feb;24(1):214–20.
39. Chou T-C, Talalay P. Mechanism of S-adenosyl-L-methionine synthesis by purified preparations of bakers' yeast. Biochemistry (Mosc). 1972 Mar 1;11(6):1065–73.
40. Yarlett N, Garofalo J, Goldberg B, Ciminelli MA, Ruggiero V, Sufrin JR, et al. S-adenosylmethionine synthetase in bloodstream *Trypanosoma brucei*. Biochim Biophys Acta. 1993 Mar 24;1181(1):68–76.
41. Sufrin JR, Lombardini JB, Alks V. Differential kinetic properties of L-2-amino-4-methylthio-cis-but-3-enoic acid, a methionine analog inhibitor of S-adenosylmethionine synthetase. Biochim Biophys Acta. 1993 Sep 3;1202(1):87–91.
42. Lindermayr C, Saalbach G, Bahnweg G, Durner J. Differential inhibition of *Arabidopsis* methionine adenosyltransferases by protein S-nitrosylation. J Biol Chem. 2006 Feb 17;281(7):4285–91.
43. Reguera RM, Balaña-Fouce R, Pérez-Pertejo Y, Fernández FJ, García-Estrada C, Cubría JC, et al. Cloning expression and characterization of methionine adenosyltransferase in Leishmania infantum promastigotes. J Biol Chem. 2002 Feb 1;277(5):3158–67.

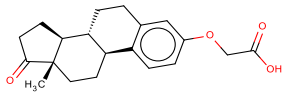
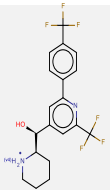
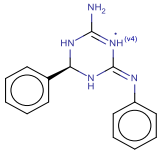
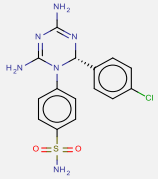
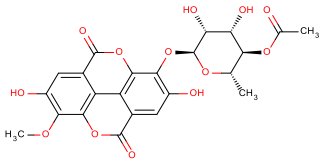
Table 7.1.7 - hits from Virtual Screen 1, listed according to predicted binding affinity.

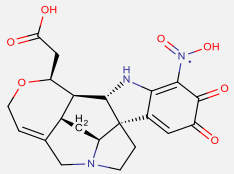
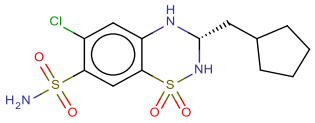
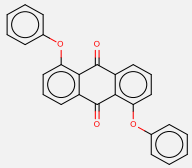
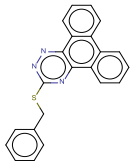
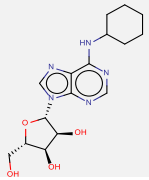
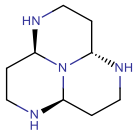
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-11.88, 1.96	ZINC01673377		296.29	1.72	181.42	10 , 4	2	✓		
-11.6, 3.14	ZINC19362650		357.56	0.73	7.68	2 , 1	2	✓		
-11.43, 4.19	ZINC01755448		398.45	6.87	52.6	2 , 0	6		✓	✓
-11.42, 4.26	apigenin-7-glucoside		432.38	0.44	166.14	10 , 6	4		✓	✓
-11.34, 4.87	daphneside		502.42	-3.36	225.06	13 , 8	6			✓

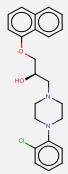
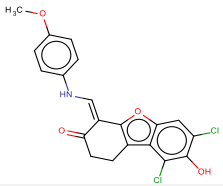
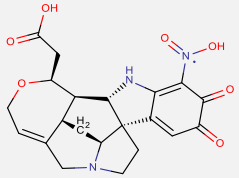
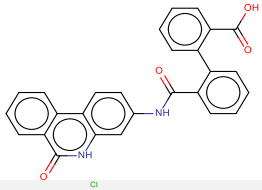
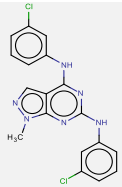
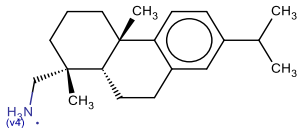
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-11.28, 5.39	ZINC19325791		182.27	0.37	39.33	4 , 3	0	✓		
-11, 8.65	lawsoniaside		500.45	-2.48	218.99	13 , 9	6		✓	✓
-10.93, 9.73	ZINC19325788		182.27	0.37	39.33	4 , 3	0	✓	✓	✓
-10.77, 12.75	ZINC19325788		182.27	0.37	39.33	4 , 3	0	✓		
-10.76, 12.97	ZINC19325794		182.27	0.37	39.33	4 , 3	0	✓	✓	✓
-10.75, 13.19	ZINC26730911		428.42	-4.61	152.45	10 , 3	3	✓	✓	✓

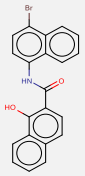
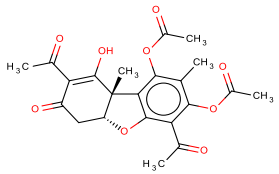
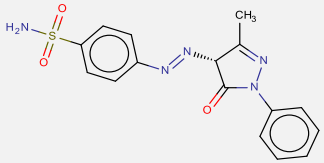
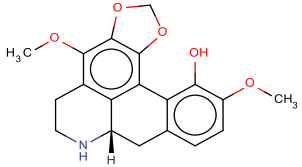
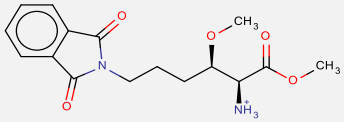
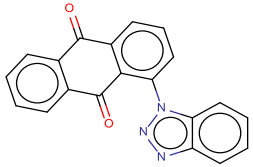
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-10.71, 14.11	ZINC01729523		353.87	0.28	49.73	3 , 2	3	✓		
-10.63, 16.15	ZINC19325794		184.28	-8.00	48.49	4 , 3	0	✓		
-10.59, 17.28	ZINC19325794		184.28	-8.00	48.49	4 , 3	0	✓		
-10.57, 17.87	ZINC18189380		265.31	3.05	74.8	5 , 3	2	✓		
-10.5, 20.11	brusatol		520.53	0.24	165.89	8 , 3	5		✓	✓
-10.42, 23.02	ZINC19325788		184.28	-8.00	48.49	4 , 3	0	✓		

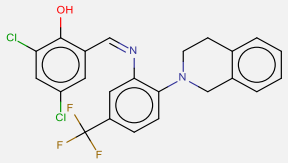
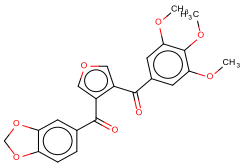
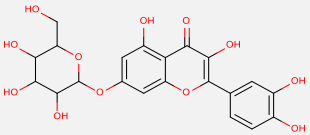
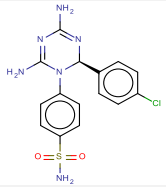
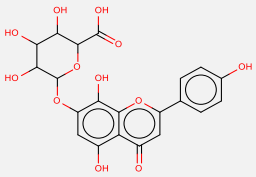
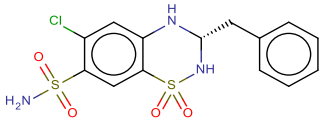
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-10.39, 24.21	luteolin-3'-O-glucoside		448.38	0.14	186.37	11 , 7	4		✓	✓
-10.36, 25.47	ZINC26730911		428.42	-4.61	152.45	10 , 3	3	✓	✓	
-10.35, 25.90	ZINC01481770		292.39	-1.08	24.67	2 , 2	0	✓		
-10.27, 29.65	ZINC01568966		407.87	5.83	92.05	4 , 2	4		✓	✓
-10.27, 29.65	ZINC01729525		353.87	0.28	49.73	3 , 2	3	✓		
-10.23, 31.72	CA60		326.31	1.97	101.38	6 , 2	3	✓	✓	✓

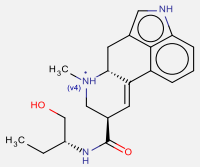
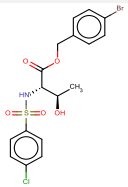
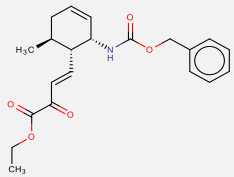
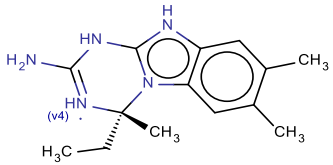
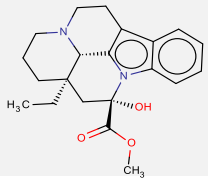
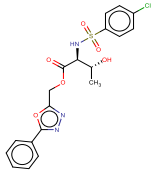
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-10.23, 31.72	ZINC04428843		328.40	3.93	63.6	4 , 1	3	✓	✓	
-10.22, 32.26	ZINC13143009		405.36	0.52	49.73	3 , 2	5	✓		
-10.2, 33.37	ZINC18189380		266.32	1.95	81.65	5 , 4	2	✓		
-10.19, 33.93	ZINC04824902		378.84	1.77	140.16	7 , 3	3	✓		✓
-10.18, 34.51	3-O-Methylellagic acid 3'-O-α-4''-O-acetylrhamnopyranoside		504.40	0.38	187.51	10 , 4	5		✓	✓

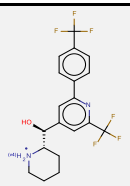
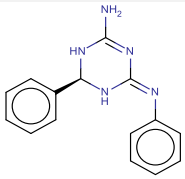
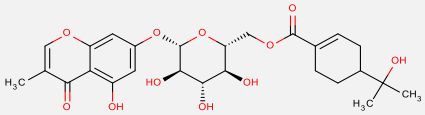
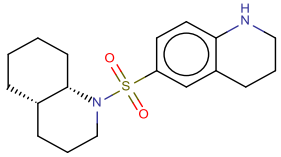
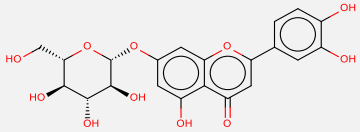
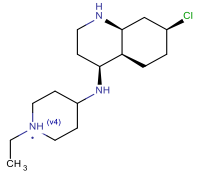
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-10.18, 34.51	ZINC19701771		428.42	-4.61	152.45	10 , 3	3	✓	✓	✓
-10.13, 37.55	ZINC03647770		379.88	1.45	118.36	5 , 3	3	✓		
-10.13, 37.55	ZINC03875548		392.40	5.92	52.6	2 , 0	4		✓	✓
-10.12, 38.19	ZINC01573829		353.44	5.74	38.67	3 , 0	3		✓	✓
-10.11, 38.84	ZINC00001761		349.38	0.01	125.55	8 , 4	4	✓	✓	✓
-10.09, 40.17	ZINC19325794		182.27	0.37	39.33	4 , 3	0	✓		

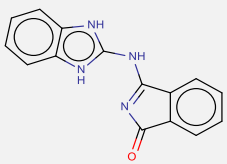
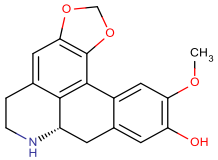
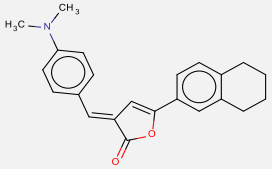
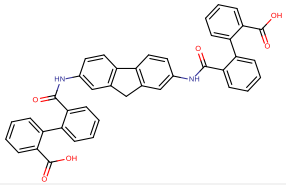
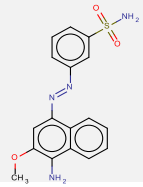
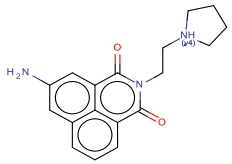
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-10.09, 40.17	ZINC29589828		396.91	4.54	35.94	4 , 1	6	✓		✓
-10.08, 40.86	ZINC05487838		404.24	4.51	71.7	4 , 2	3	✓	✓	✓
-10.04, 43.71	ZINC19701771		428.42	-4.61	152.45	10 , 3	3	✓		
-10.01, 45.98	ZINC29589888		434.44	5.14	95.5	4 , 3	4			✓
-9.99, 47.56	ZINC01718486		385.25	5.23	67.66	5 , 2	4			✓
-9.97, 49.19	ZINC03947435		286.47	1.32	27.64	1 , 1	2	✓		

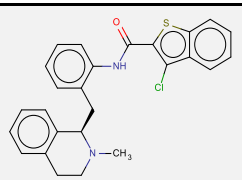
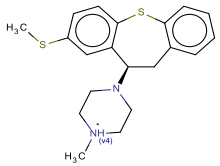
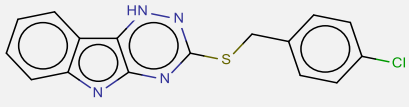
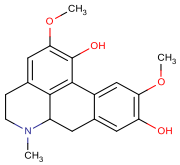
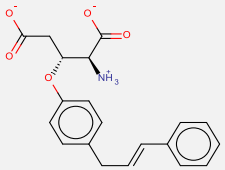
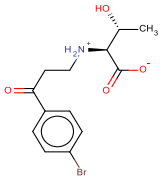
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-9.93, 52.63	ZINC01607786		392.25	5.51	49.33	2 , 2	2			
-9.92, 53.52	ZINC01563325		430.40	1.02	133.27	7 , 1	6	✓		
-9.91, 54.43	ZINC22910159		357.39	2.22	117.55	6 , 1	4	✓		
-9.9, 55.36	cassyformine		341.36	1.73	69.18	6 , 2	2	✓		
-9.9, 55.36	ZINC01708940		321.35	0.55	100.55	4 , 1	8	✓		
-9.88, 57.26	ZINC04896472		325.32	4.03	64.85	4 , 0	1	✓	✓	✓

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-9.88, 57.26	ZINC18124911		465.30	7.50	35.83	3 , 1	4			
-9.87, 58.24	ZINC01722585		410.37	3.18	93.43	7 , 0	7	✓		
-9.84, 61.26	quercetin 7-glucosyl		464.38	-0.11	206.6	12 , 8	4		✓	✓
-9.8, 65.54	ZINC01707215		378.84	1.77	140.16	7 , 3	3	✓		
-9.77, 68.94	isoscutellarin		462.36	0.46	203.44	12 , 7	4		✓	✓
-9.77, 68.94	ZINC02015922		387.86	1.42	118.36	5 , 3	3	✓		

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-9.77, 68.94	ZINC03831068		340.44	-2.52	69.56	3 , 4	4	✓		✓
-9.74, 72.52	ZINC03260391		462.74	3.46	92.7	4 , 2	7	✓	✓	
-9.7, 77.59	ZINC01590993		371.43	4.37	81.7	3 , 1	9	✓		
-9.69, 78.91	ZINC13597738		259.35	4.03	77.98	3 , 4	1	✓		
-9.68, 80.25	ZINC00608128		354.44	3.39	54.7	3 , 1	3	✓		
-9.67, 81.62	ZINC02636757		451.88	1.77	131.62	6 , 2	8	✓		✓

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-9.67, 81.62	ZINC13143008		405.36	0.52	49.73	3 , 2	5	✓		
-9.67, 81.62	ZINC18189380		265.31	2.59	74.8	5 , 3	2	✓		
-9.66, 83.01	camaldulenside		520.53	2.10	172.21	10 , 5	7		✓	✓
-9.66, 83.01	ZINC05493736		334.48	3.12	49.41	3 , 1	1	✓	✓	✓
-9.65, 84.42	luteolin-7-O-glucoside		448.38	0.14	186.37	11 , 7	4		✓	✓
-9.64, 85.86	ZINC01683345		300.89	-3.18	28.5	3 , 3	3	✓		

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-9.62, 88.81	ZINC05647206		263.27	2.65	73.04	3 , 3	1	✓		
-9.58, 95.01	actinodaphnine		311.33	1.91	59.95	5 , 2	1	✓		
-9.57, 96.63	ZINC01665801		345.43	5.45	29.54	2 , 0	3			
-9.56, 98.27	NSC128590 (CHEMBL1172557)		644.67	8.53	132.8	6 , 4	8			✓
-9.56, 98.27	ZINC18038389		356.40	2.99	120.13	6 , 2	4	✓		
-9.55, 99.94	ZINC05640620		310.37	-3.07	67.84	4 , 2	3	✓		

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-9.55, 99.94	ZINC29590277		446.99	6.73	32.34	2 , 1	4		✓	✓
-9.54, 101.65	ZINC19362651		357.56	0.73	7.68	2 , 1	2	✓		
-9.52, 105.14	ZINC06576323		326.80	4.65	54.46	3 , 1	3	✓		
-9.51, 106.92	D-isoboldine		327.37	2.78	62.16	5 , 2	2	✓		
-9.51, 106.92	ZINC00600877		354.38	0.87	117.13	5 , 1	9	✓		✓
-9.51, 106.92	ZINC04141542 (BPOB)		330.17	-1.01	94.04	4 , 2	7	✓		

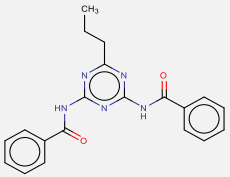
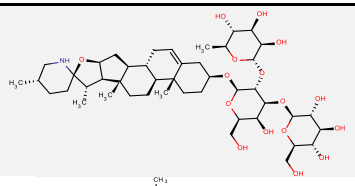
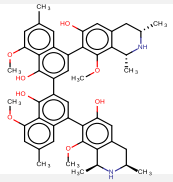
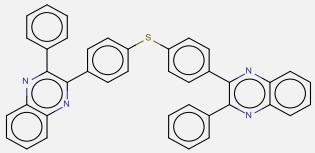
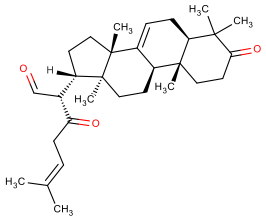
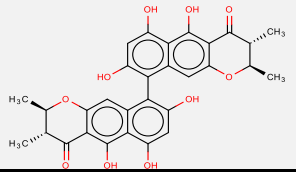
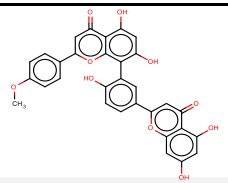
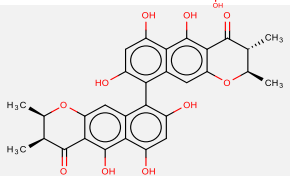
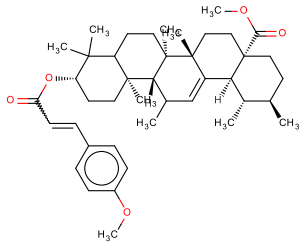
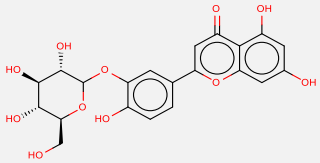
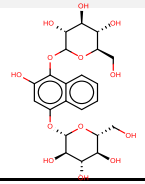
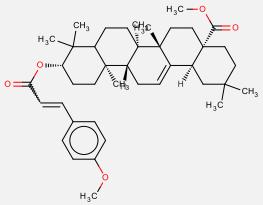
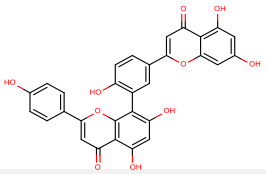
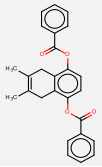
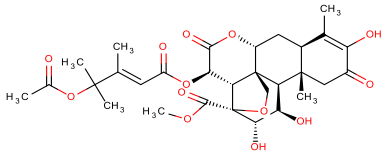
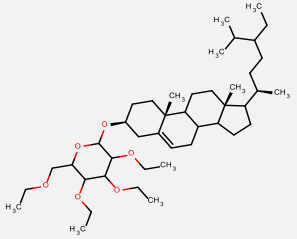
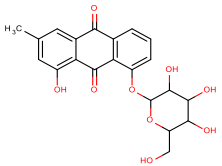
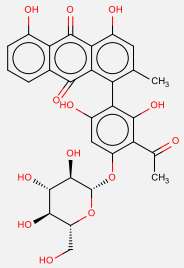
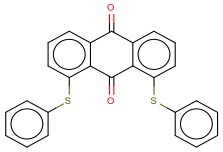
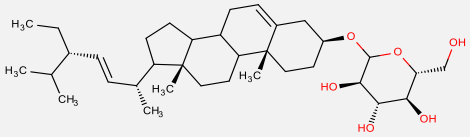
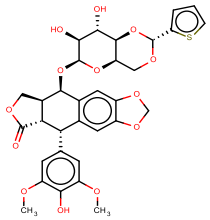
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS3?	Hit in VS4?
-9.5, 108.74	ZINC01584497		361.40	4.75	96.87	5 , 2	6	✓		✓

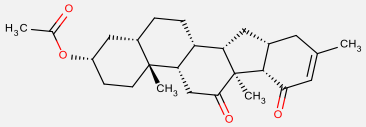
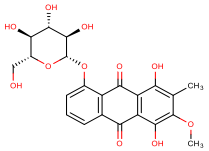
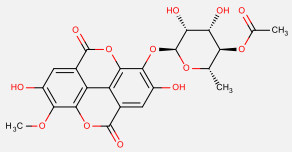
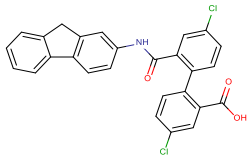
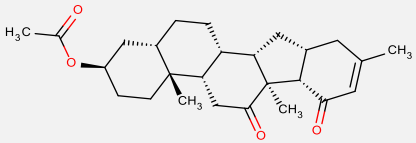
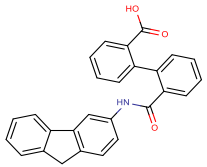
Table 7.1.8 - hits from Virtual Screen 3, listed according to predicted binding energy.

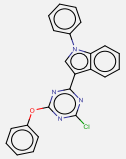
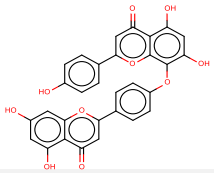
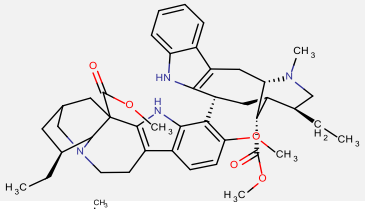
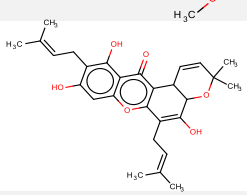
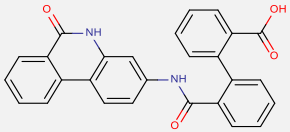
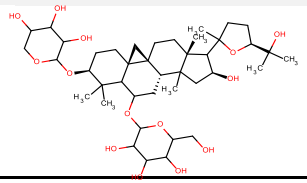
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-15.03, .01	α-solamargine (solasonine)		884.06	0.34	258.71	17 , 10	8			
-13.62, .10	ancistrogriffithine A		784.94	7.21	141.9	10 , 6	5			
-13.62, .10	ZINC05492794		594.73	10.95	51.56	4 , 0	6			✓
-13.11, .25	3,23-dioxotirucalla-7,24-dien-21-al		452.67	6.42	51.21	3 , 0	5			✓
-12.72, .47	ZINC17465983		546.52	5.67	173.98	10 , 6	0			✓

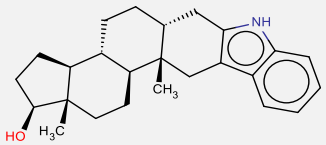
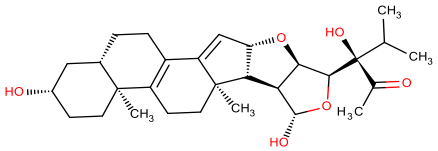
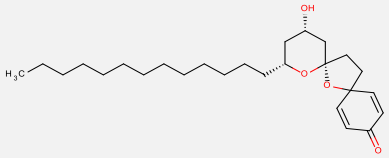
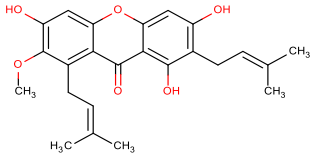
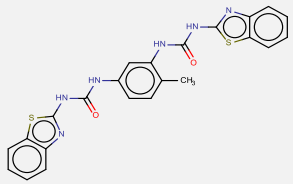
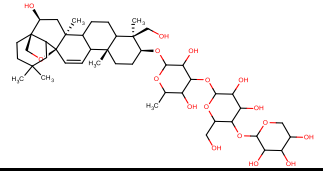
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-12.7, .49	podocarpusflavone A		552.48	5.23	162.98	10 , 5	3			✓
-12.56, .62	ZINC17465979		546.52	5.67	173.98	10 , 6	0			✓
-12.28, 1.00	Cis-*p*-methoxycinnamoyloxyursolicacid methyl ester		644.92	9.89	61.83	3 , 0	7		✓	✓
-12.28, 1.00	luteolin-3'-O-glucoside		448.38	0.14	186.37	11 , 7	4		✓	✓
-12.2, 1.14	lawsoniaside		500.45	-2.48	218.99	13 , 9	6		✓	✓

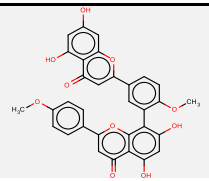
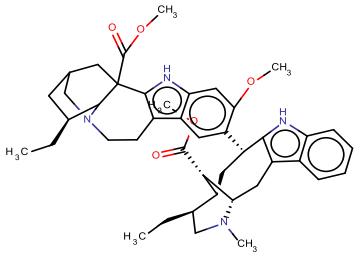
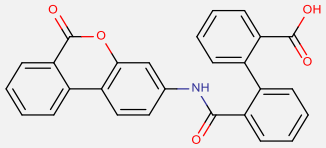
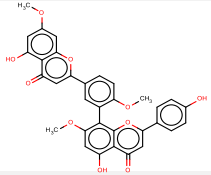
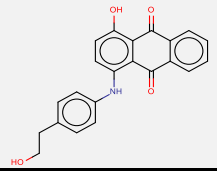
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-12.28, 1.00	Cis-*p*-methoxycinnamoyloxyoleanolicacid methyl ester		630.90	9.62	61.83	3 , 0	7			✓
-12.13, 1.28	amentoflavone		538.46	5.09	173.98	10 , 6	2			✓
-12.07, 1.42	ZINC01755448		398.45	6.87	52.6	2 , 0	6		✓	✓
-11.87, 1.99	bruceantinol		606.61	0.10	192.19	9 , 3	8			✓
-11.87, 1.99	tetra-acetyldaucosterol		689.06	10.07	55.38	6 , 0	17			✓

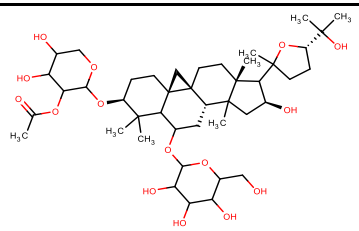
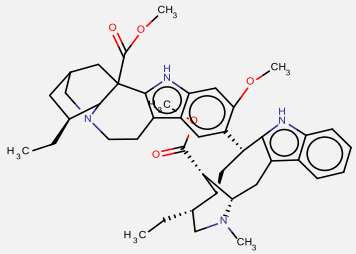
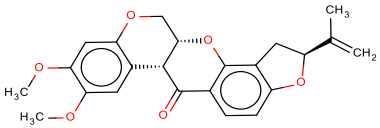
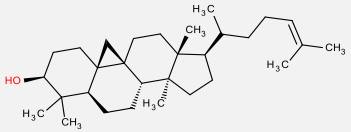
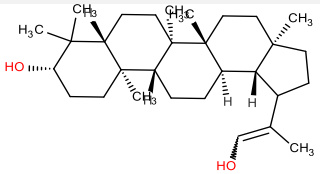
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-11.86, 2.03	chrysophanol-8-glucoside		416.38	1.21	153.75	9 , 5	3	✓		✓
-11.81, 2.20	4'-O-demethylknipholone-4'-O-B-d-glucopyranoside		582.51	2.80	231.51	13 , 8	4			✓
-11.73, 2.52	ZINC05124931		424.53	7.30	34.14	2 , 0	4			✓
-11.71, 2.61	stigmasterol-D-glucoside		574.83	5.71	99.38	6 , 4	8			✓
-11.68, 2.74	ZINC11616856		656.65	2.78	160.83	12 , 3	6			

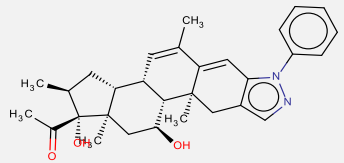
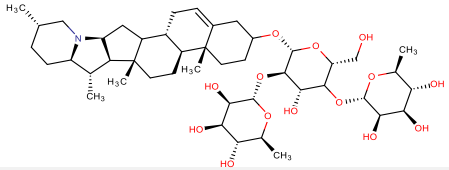
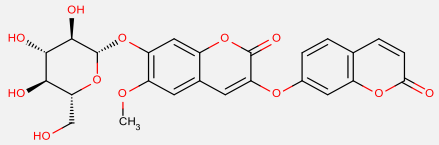
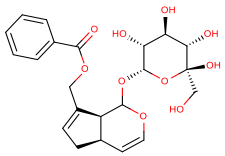
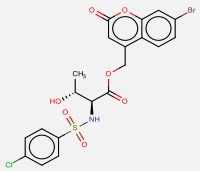
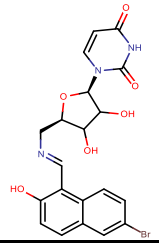
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-11.63, 2.99	ZINC12671904		412.56	4.53	60.44	3 , 0	2	✓		✓
-11.61, 3.09	physcion 1-β-D-glucopyranoside		462.40	1.40	183.21	11 , 6	4			✓
-11.49, 3.78	3-O-Methylellagic acid 3'-O-α-4''-O-acetylrhamnopyranoside		504.40	0.38	187.51	10 , 4	5		✓	✓
-11.48, 3.85	NSC132249 (CHEMBL1173812)		474.34	7.34	66.4	3 , 2	4			✓
-11.37, 4.63	ZINC12671898		412.56	4.53	60.44	3 , 0	2	✓		✓
-11.36, 4.71	NSC128594 (CHEMBL1173451)		405.44	6.14	66.4	3 , 2	4			✓

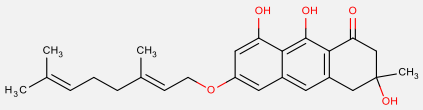
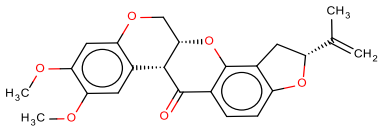
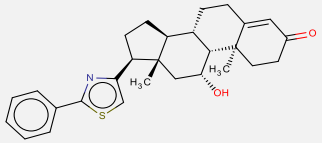
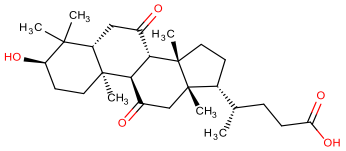
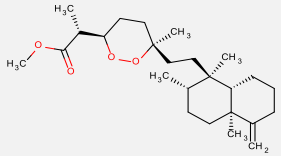
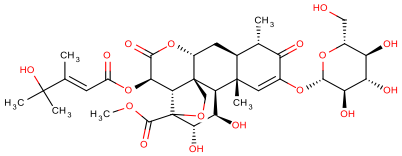
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-11.36, 4.71	ZINC01726776		398.84	7.10	52.83	3 , 0	4			
-11.35, 4.79	lanaroflavone		538.46	5.24	162.98	9 , 5	4			✓
-11.3, 5.21	tabernaelegantine A		706.91	6.53	99.89	5 , 2	8			✓
-11.3, 5.21	tovophyllin A		464.55	5.36	96.22	6 , 3	4			✓
-11.27, 5.48	NSC127133 (CHEMBL1172911)		434.44	5.14	95.5	4 , 3	4			✓
-11.24, 5.77	astrogaloside III		784.97	-0.31	228.22	14 , 9	7			✓

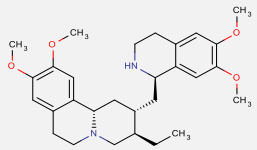
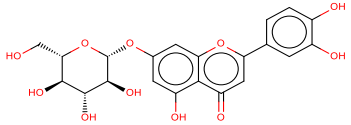
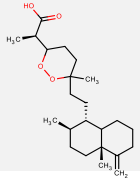
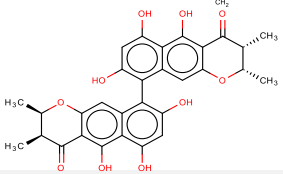
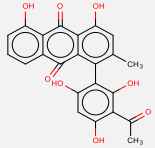
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-11.23, 5.87	ZINC13154307		363.54	5.21	36.02	1 , 2	0			✓
-11.22, 5.97	vernoguinosterol		486.64	2.84	96.22	6 , 3	3	✓		✓
-11.19, 6.28	aculeatin D		418.61	7.02	55.76	4 , 1	12			✓
-11.19, 6.28	mangostin		410.46	6.00	96.22	5 , 3	5			
-11.18, 6.38	ZINC01572309		474.56	6.43	108.04	4 , 4	4			✓
-11.17, 6.49	buddlejasaponin III		899.07	-0.04	266.91	17 , 10	8			

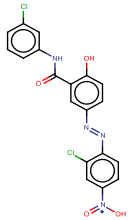
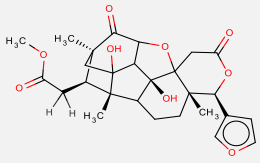
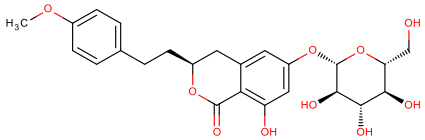
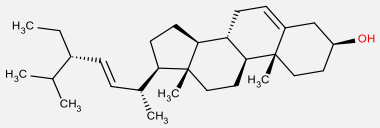
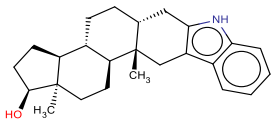
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-11.15, 6.71	isoginkgetin		566.51	5.38	151.98	10 , 4	4		✓	✓
-11.11, 7.18	tabernaelegantine B		706.91	6.53	99.89	5 , 2	8			
-11.05, 7.95	NSC128604 (CHEMBL1173036)		435.43	5.32	92.7	4 , 2	4			✓
-11.05, 7.95	ginkgetin		580.54	5.53	140.98	10 , 3	5			✓
-11.04, 8.08	ZINC05124960		359.37	5.53	86.63	5 , 3	4			✓

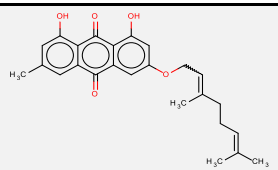
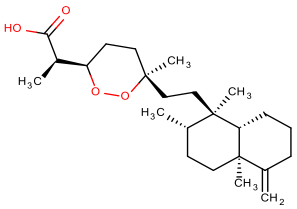
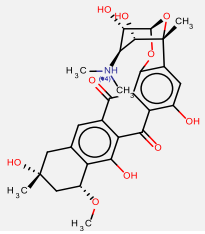
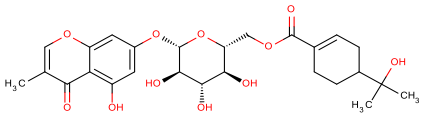
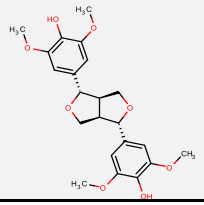
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-11.02, 8.36	astrogaloside II		827.01	0.13	234.29	14 , 8	9			
-11.01, 8.50	tabernaelegantine D		706.91	6.53	99.89	5 , 2	8			
-11.01, 8.50	ZINC01530836		394.42	3.32	63.22	6 , 0	3	✓		
-10.99, 8.79	cycloartenol		426.72	7.55	20.23	1 , 1	4			✓
-10.99, 8.79	hennadiol		442.72	6.86	40.46	2 , 2	1			✓

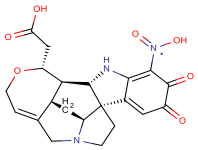
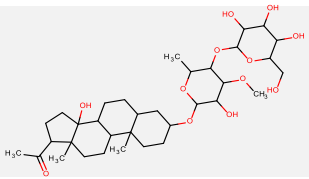
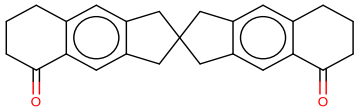
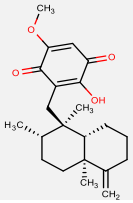
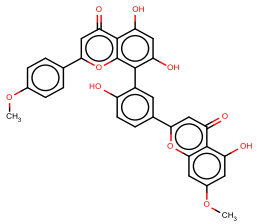
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.99, 8.79	ZINC17353912		472.62	4.00	75.35	4 , 2	2	✓		✓
-10.97, 9.10	α-chaconine		852.06	0.61	220.46	15 , 8	7			
-10.97, 9.10	daphnorin		514.44	0.42	170.44	10 , 4	6			✓
-10.94, 9.57	melampyroside		450.44	0.28	155.14	9 , 5	7	✓		✓
-10.92, 9.90	ZINC03417739		530.77	3.06	119	5 , 2	7			✓
-10.86, 10.95	6-Bromo-2-hydroxy-naphthalene-1-carbaldehyde O-[5-(2,4-dioxo-3,4-dihydro-2H-pyrimidin-1-yl)-3,4-dihydroxy-tetrahydro-furan-2-ylmethyl]-oxime		476.28	1.48	131.69	7 , 4	4	✓		

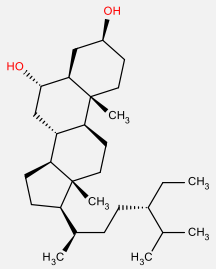
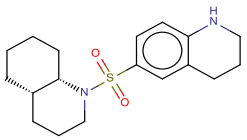
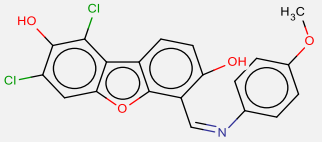
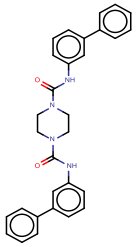
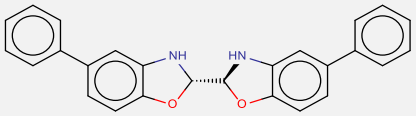
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.86, 10.95	vismione D		410.50	5.23	86.99	5 , 3	6			✓
-10.86, 10.95	ZINC01705443		394.42	3.32	63.22	6 , 0	3	✓		
-10.86, 10.95	ZINC13099048		447.63	5.52	50.19	3 , 1	2			✓
-10.84, 11.33	ZINC12670914		446.62	4.22	91.67	5 , 2	4	✓		✓
-10.83, 11.52	sigmosceptrellin A methyl ester		406.60	6.27	44.76	3 , 0	6			✓
-10.83, 11.52	yadanzioside C		726.72	-2.35	265.27	14 , 7	9			✓

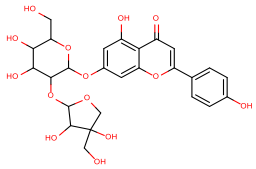
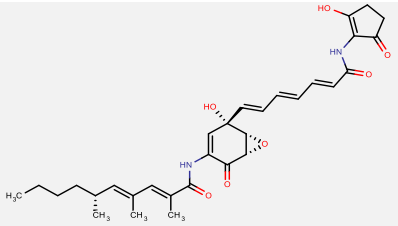
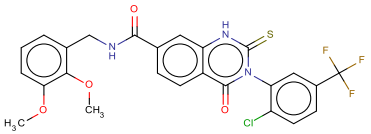
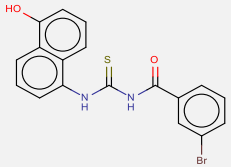
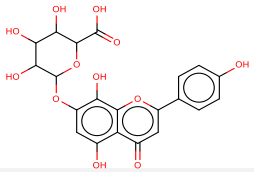
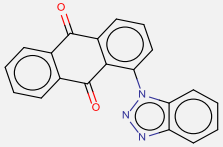
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.79, 12.33	emetine		480.64	4.49	52.19	6 , 1	7	✓		✓
-10.77, 12.75	luteolin-7-O-glucoside		448.38	0.14	186.37	11 , 7	4		✓	✓
-10.73, 13.64	sigmosceptrellin B		378.55	5.82	55.76	4 , 1	5			✓
-10.72, 13.87	ZINC05462674		546.52	5.67	173.98	10 , 6	0			
-10.71, 14.11	G4'-O-demethylknipholone		420.37	5.72	152.36	8 , 5	1			✓

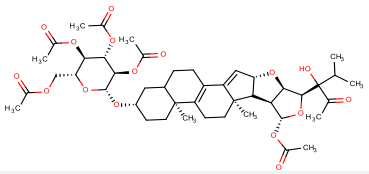
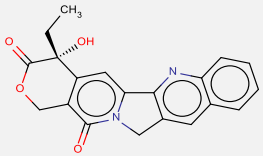
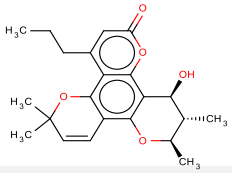
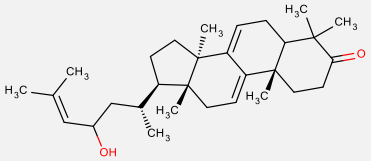
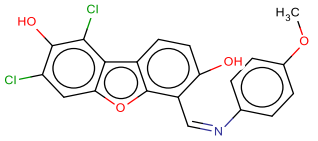
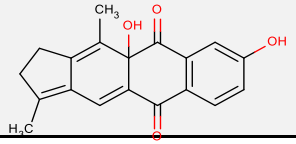
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.71, 14.11	ZINC17995347		432.24	4.90	130.56	7 , 3	5	✓		
-10.7, 14.35	6-dehydroxykhayanolide E		500.54	1.27	132.5	6 , 2	4			✓
-10.66, 15.35	agrimonolide 6-O-β-D-glucopyranoside		476.47	1.95	155.14	9 , 5	7	✓		✓
-10.66, 15.35	stigmasterol		412.69	7.48	20.23	1 , 1	5			✓
-10.64, 15.88	ZINC13154306		363.54	5.21	36.02	1 , 2	0			

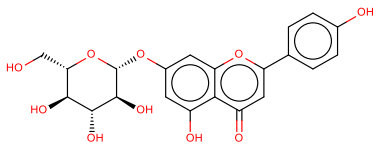
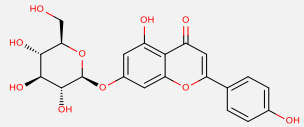
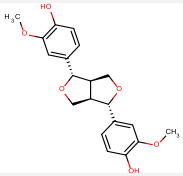
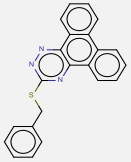
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.62, 16.42	3-geranylemodin		406.47	6.99	83.83	5 , 2	6		✓	✓
-10.59, 17.28	sigmosceptrellin A		392.57	6.12	55.76	4 , 1	5			✓
-10.58, 17.57	ZINC11681166		542.55	-3.47	167.42	11 , 6	2			✓
-10.57, 17.87	camaldulenside		520.53	2.10	172.21	10 , 5	7		✓	✓
-10.56, 18.17	(+)-syringaresinol		418.44	1.96	95.84	8 , 2	6	✓		✓

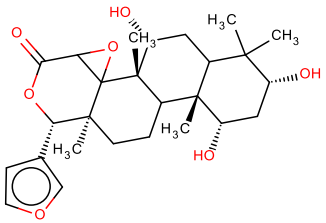
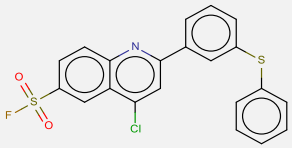
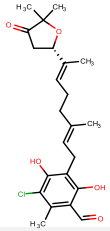
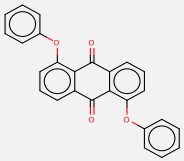
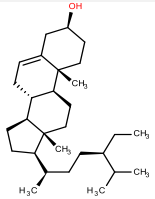
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.54, 18.80	ZINC26730911		428.42	-4.61	152.45	10 , 3	3	✓	✓	✓
-10.53, 19.12	caratuberside A		656.80	1.05	184.6	12 , 6	7			✓
-10.53, 19.12	ZINC01855333		356.46	5.37	34.14	2 , 0	0			✓
-10.52, 19.44	ilimaquinone		358.47	4.37	63.6	4 , 1	3	✓		✓
-10.52, 19.44	podocarpusflavone B		566.51	5.38	151.98	10 , 4	4			✓

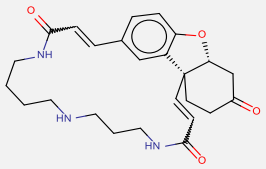
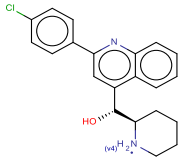
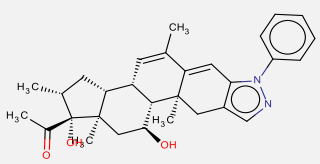
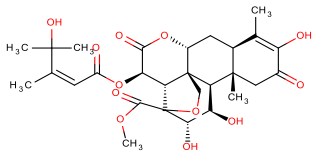
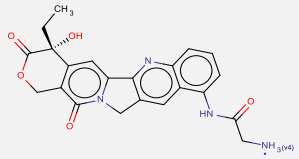
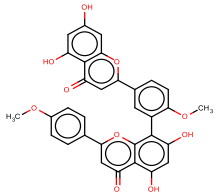
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.51, 19.77	5α-poriferastane-3β,6α-diol		432.72	6.94	40.46	2 , 2	6			✓
-10.51, 19.77	ZINC05493736		334.48	3.12	49.41	3 , 1	1	✓	✓	✓
-10.51, 19.77	ZINC17949075		402.23	5.47	75.19	4 , 2	3		✓	✓
-10.49, 20.45	ZINC04783229		476.57	5.60	64.68	2 , 2	4			✓
-10.48, 20.80	ZINC01736227		392.45	6.01	42.52	4 , 2	3			✓

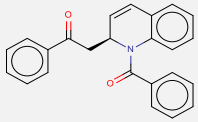
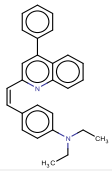
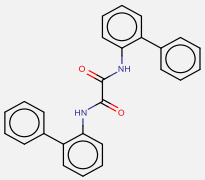
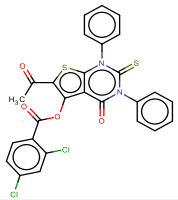
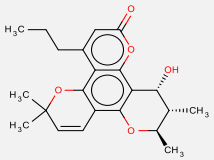
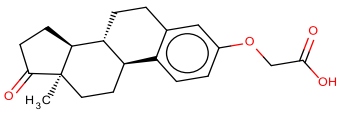
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.46, 21.51	apiin		564.49	-0.84	225.06	14 , 8	7			✓
-10.45, 21.88	manumycin A		550.64	3.06	145.33	7 , 4	12			✓
-10.45, 21.88	Qc2		549.95	5.48	79.9	4 , 2	6			✓
-10.45, 21.88	ZINC05580600		401.28	5.06	61.36	2 , 3	2			
-10.44, 22.25	isoscutellarin		462.36	0.46	203.44	12 , 7	4		✓	✓
-10.44, 22.25	ZINC04896472		325.32	4.03	64.85	4 , 0	1	✓	✓	✓

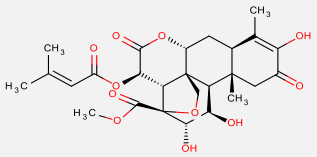
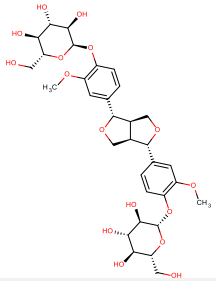
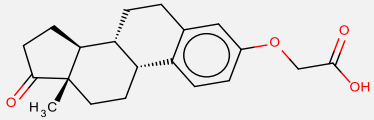
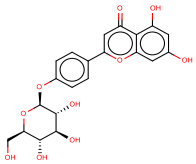
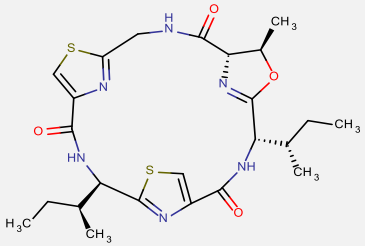
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.43, 22.63	vernoguinoside		858.96	3.27	205.72	11 , 1	16			
-10.43, 22.63	ZINC00001087		348.35	1.22	79.73	4 , 1	1	✓		✓
-10.43, 22.63	ZINC01645454		370.44	3.79	64.99	4 , 1	2	✓		
-10.42, 23.02	23-Hydroxy-5α-lanosta-7,9(11),24-triene-3-one		438.69	6.68	37.3	2 , 1	4			✓
-10.42, 23.02	ZINC17949075		402.23	5.47	75.19	4 , 2	3			
-10.41, 23.41	kigelinol		308.33	1.83	74.6	4 , 2	0	✓		✓

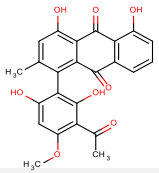
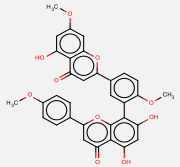
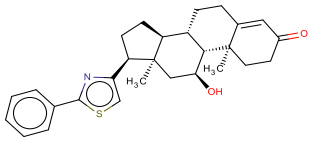
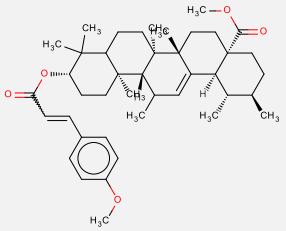
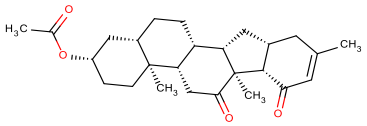
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.41, 23.41	cosmossin (acacetin-7-O-glucoside)		432.38	0.44	166.14	10 , 6	4			✓
-10.4, 23.81	apigenin-7-glucoside		432.38	0.44	166.14	10 , 6	4		✓	✓
-10.4, 23.81	pinoresinol		358.39	2.28	77.38	6 , 2	4	✓		
-10.4, 23.81	ZINC01573829		353.44	5.74	38.67	3 , 0	3		✓	✓

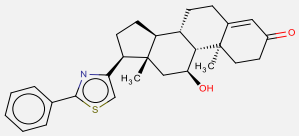
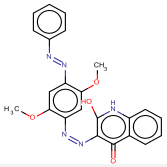
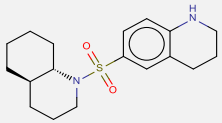
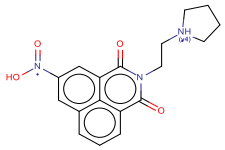
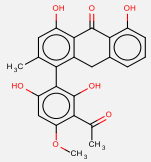
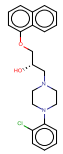
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.37, 25.04	1α,3α,7α-trideacetylkhivorin		460.56	1.99	112.66	5 , 3	1	✓		✓
-10.36, 25.47	ZINC17465958		429.92	6.62	47.03	3 , 0	4			✓
-10.35, 25.90	ascofuranone		420.93	6.40	83.83	5 , 2	7			✓
-10.35, 25.90	ZINC03875548		392.40	5.92	52.6	2 , 0	4		✓	✓
-10.34, 26.34	B-sitosterol		414.71	7.84	20.23	1 , 1	6			✓

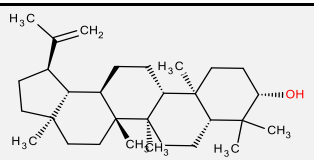
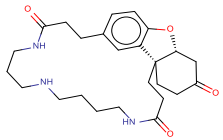
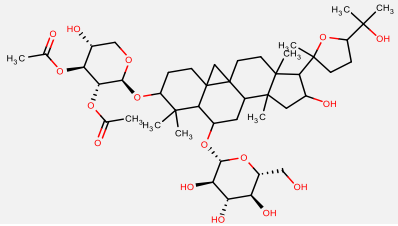
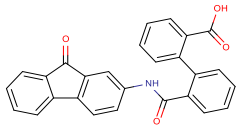
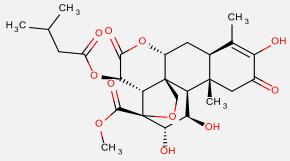
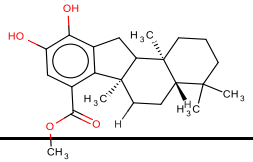
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.34, 26.34	lunaridine		437.53	1.16	96.53	5 , 3	0	✓		
-10.32, 27.25	ZINC01729524		353.87	0.28	49.73	3 , 2	3	✓		
-10.32, 27.25	ZINC17353911		472.62	4.00	75.35	4 , 2	2	✓		✓
-10.31, 27.71	bruceine C		564.58	-0.35	186.12	9 , 4	6			✓
-10.31, 27.71	ZINC13000556		421.43	-5.26	136.47	6 , 3	3	✓		✓
-10.29, 28.66	bilobetin		566.51	5.38	151.98	10 , 4	4			✓

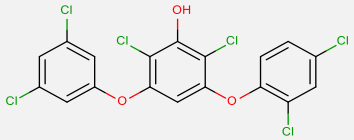
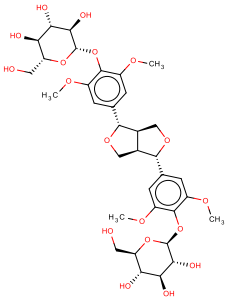
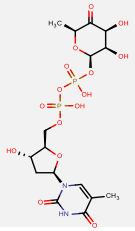
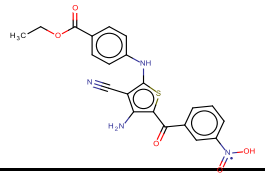
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.27, 29.65	ZINC01705925		353.41	4.82	37.38	2 , 0	4	✓		
-10.27, 29.65	ZINC08581460		378.51	7.17	16.13	2 , 0	6			✓
-10.25, 30.67	ZINC01577889		392.45	5.90	58.2	2 , 2	5			
-10.25, 30.67	ZINC01644254		567.46	7.55	66.92	3 , 0	6			✓
-10.24, 31.19	ZINC00600322		370.44	3.79	64.99	4 , 1	2	✓		
-10.24, 31.19	ZINC13099027		328.40	3.93	63.6	4 , 1	3	✓		

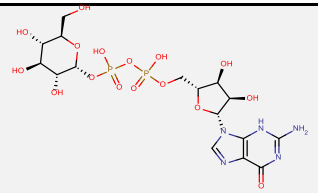
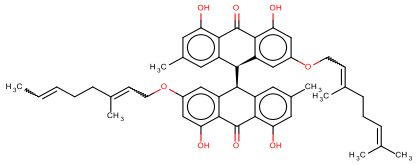
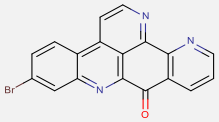
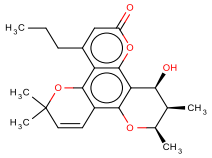
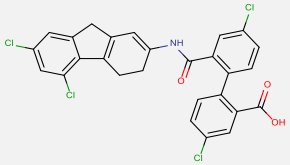
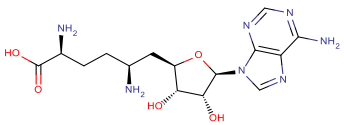
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.23, 31.72	brusatol		520.53	0.24	165.89	8 , 3	5		✓	✓
-10.22, 32.26	pinoresinol di-O-β-D-glucopyranoside		682.67	-2.26	235.68	16 , 8	10			✓
-10.21, 32.81	ZINC13099025		328.40	3.93	63.6	4 , 1	3	✓		✓
-10.2, 33.37	apigenin-4'-glucoside		432.38	0.44	166.14	10 , 6	4			✓
-10.19, 33.93	aerucyclamide B		532.68	2.81	134.67	6 , 3	4			

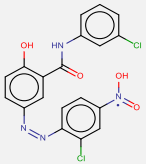
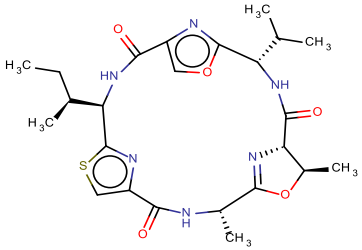
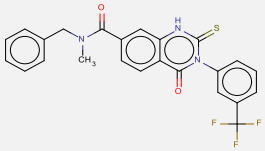
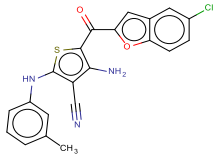
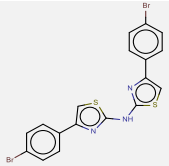
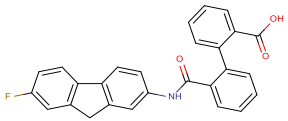
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.18, 34.51	knipholone		434.39	5.22	141.36	8 , 4	2			✓
-10.18, 34.51	sciadopitysin		580.54	5.53	140.98	10 , 3	5			✓
-10.18, 34.51	ZINC13099050		447.63	5.52	50.19	3 , 1	2			✓
-10.16, 35.70	*trans-p*-methoxycinnamoyloxyursolicacid methyl ester		644.92	9.89	61.83	3 , 0	7			✓
-10.16, 35.70	ZINC12671893		412.56	4.53	60.44	3 , 0	2	✓		✓

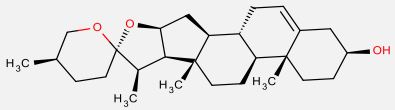
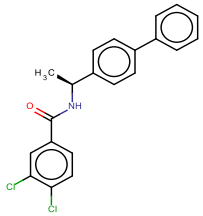
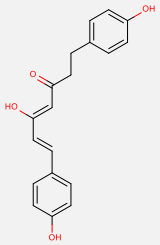
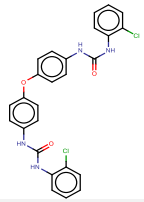
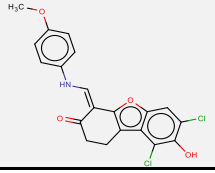
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.16, 35.70	ZINC13099051		447.63	5.52	50.19	3 , 1	2			✓
-10.16, 35.70	ZINC18169477		429.43	6.21	117.23	9 , 2	6			✓
-10.15, 36.30	ZINC01686969		334.48	3.12	49.41	3 , 1	1	✓		
-10.15, 36.30	ZINC04217305		341.36	-3.47	98.33	6 , 2	4	✓		✓
-10.14, 36.92	knipholone anthrone		420.41	5.85	124.29	7 , 4	2			✓
-10.14, 36.92	ZINC29589829		396.91	4.54	35.94	4 , 1	6	✓		

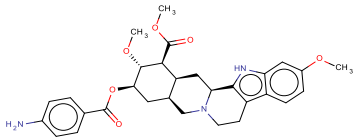
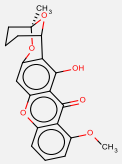
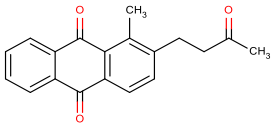
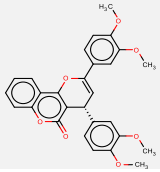
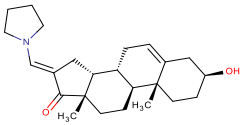
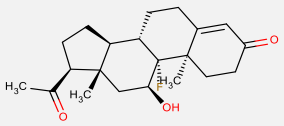
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.13, 37.55	lupeol		426.72	7.45	20.23	1 , 1	1			
-10.13, 37.55	tetrahydrolunarine		441.56	1.08	96.53	5 , 3	0	✓		✓
-10.12, 38.19	astrogaloside I		869.04	0.57	240.36	14 , 7	11			
-10.11, 38.84	NSC67692 (CHEMBL1172978)		419.43	5.50	83.47	4 , 2	4			✓
-10.11, 38.84	bruceine A		522.54	0.05	165.89	8 , 3	6			✓
-10.1, 39.50	pelorol		372.50	5.56	66.76	3 , 2	2			

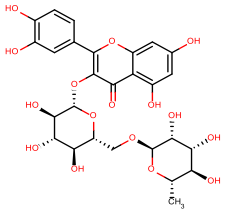
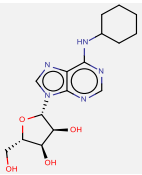
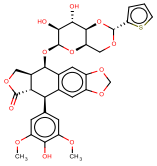
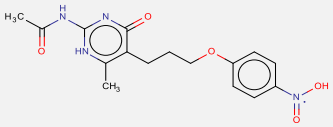
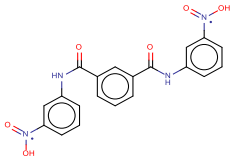
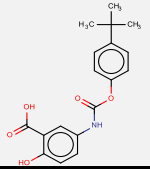
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10.05, 42.98	ambigol C		484.97	8.29	38.69	1 , 1	4			
-10.03, 44.45	Syringaresinol di-O-β-D-glucopyranoside**		742.72	-2.57	254.14	18 , 8	12			
-10.02, 45.21	dTDP-6-deoxy-L-lyxo-4-hexulose		546.31	-2.04	247.92	12 , 6	8			
-10.02, 45.21	ZINC01735469		437.45	3.51	161.72	8 , 3	8	✓		

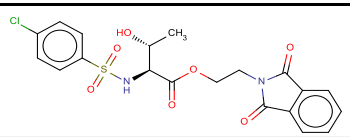
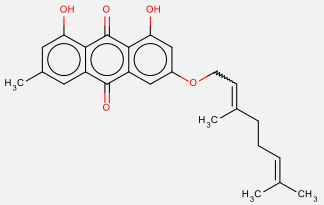
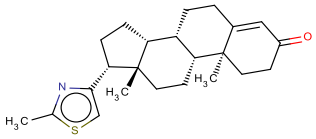
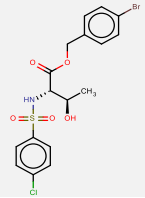
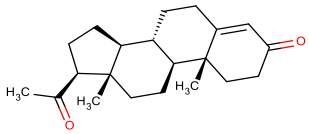
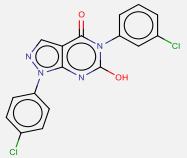
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-10, 46.76	GDP-L-glucose		605.34	-4.77	327.43	17 , 10	9			✓
-10, 46.76	bianthrone A1		768.93	14.22	133.52	8 , 4	13			
-10, 46.76	2-bromoascicidemin		362.18	3.74	55.74	4 , 0	0	✓		
-10, 46.76	ZINC05536814		370.44	3.79	64.99	4 , 1	2	✓		
-9.99, 47.56	CHEMBL1172759		545.24	7.24	66.4	3 , 2	4			✓
-9.99, 47.56	sinefungin		381.39	-5.14	208.65	11 , 6	7			✓

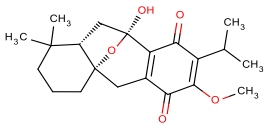
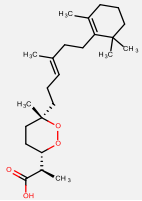
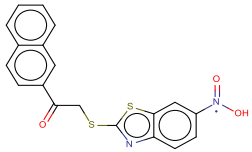
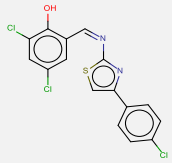
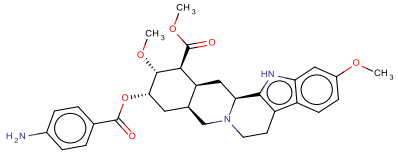
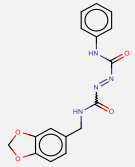
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.97, 49.19	ZINC17995347		432.24	4.90	130.56	7 , 3	5	✓		
-9.96, 50.03	aerucyclamide C		516.61	2.22	147.81	6 , 3	3			
-9.96, 50.03	Qc6		469.48	5.42	52.65	2 , 1	5			
-9.96, 50.03	ZINC01568966		407.87	5.83	92.05	4 , 2	4		✓	✓
-9.96, 50.03	ZINC13154298		493.24	7.70	37.81	3 , 1	4			✓
-9.94, 51.75	NSC128598 (CHEMBL1173453)		423.44	6.28	66.4	3 , 2	4			

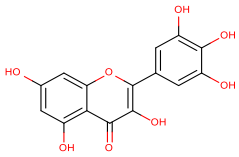
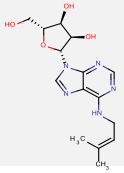
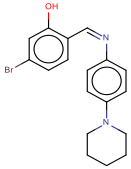
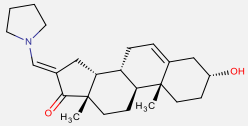
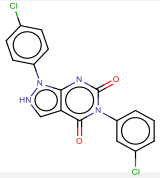
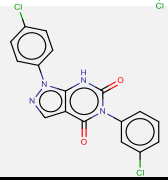
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.94, 51.75	nitogenin (diosgenin)		414.62	4.93	38.69	3 , 1	0	✓		✓
-9.94, 51.75	ZINC01735767		370.27	6.04	29.1	1 , 1	4			✓
-9.93, 52.63	(4Z,6E)-5-hydroxy-1,7-bis(4-hydroxyphenyl)hepta-4,6-dien-3-one		310.34	3.99	77.76	4 , 3	6	✓		✓
-9.93, 52.63	ZINC01639633		507.37	6.97	91.49	2 , 4	6			
-9.92, 53.52	ZINC05487838		404.24	4.51	71.7	4 , 2	3	✓	✓	✓

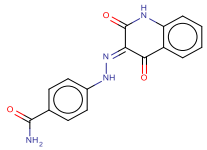
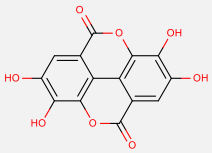
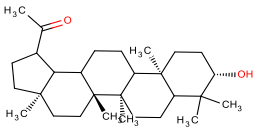
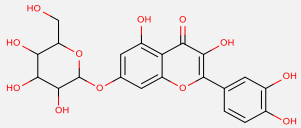
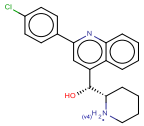
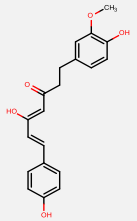
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.92, 53.52	ZINC17287336		533.62	3.18	116.11	6 , 2	7			
-9.91, 54.43	chaetoxanthone B		354.35	3.95	74.22	5 , 1	1	✓		
-9.91, 54.43	sterequinone F		292.33	3.84	51.21	3 , 0	3	✓		
-9.91, 54.43	ZINC02034999		472.49	3.75	72.45	6 , 0	6	✓		
-9.91, 54.43	ZINC13597348		369.54	3.81	40.54	3 , 1	1	✓		✓
-9.91, 54.43	ZINC17147431		348.45	2.88	54.37	3 , 1	1	✓		

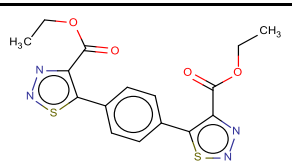
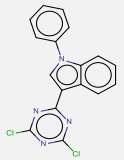
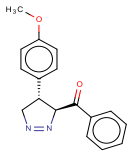
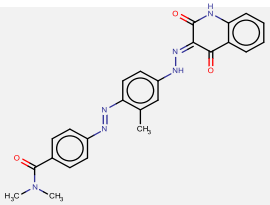
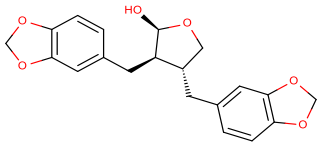
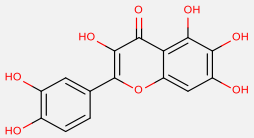
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.9, 55.36	rutin		610.52	-0.87	265.52	16 , 10	6			✓
-9.9, 55.36	ZINC00001761		349.38	0.01	125.55	8 , 4	4	✓	✓	✓
-9.9, 55.36	ZINC11616857		656.65	2.78	160.83	12 , 3	6			
-9.9, 55.36	ZINC13597732		347.35	-0.29	136.3	8 , 3	7	✓		
-9.89, 56.30	ZINC01738764		408.36	1.20	171.22	8 , 4	6	✓		
-9.89, 56.30	ZINC01748097		329.35	5.04	95.86	5 , 3	5			

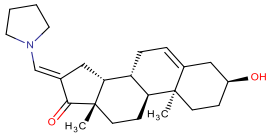
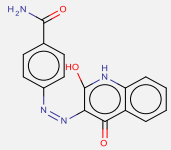
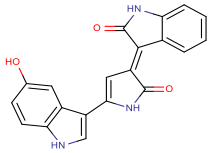
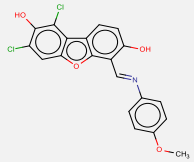
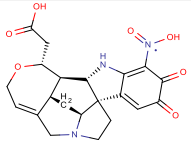
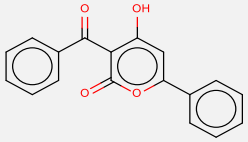
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.89, 56.30	ZINC03297153		466.89	1.71	130.08	6 , 2	8	✓		
-9.88, 57.26	3-geranyloxy-6-methyl-1,8-dihydroxyanthraquinone		406.47	6.99	83.83	5 , 2	6			✓
-9.88, 57.26	ZINC17375647		369.56	4.93	29.96	2 , 0	1	✓		✓
-9.87, 58.24	ZINC03260391		462.74	3.46	92.7	4 , 2	7	✓	✓	
-9.86, 59.23	progesterone		314.46	4.15	34.14	2 , 0	1	✓		
-9.86, 59.23	ZINC06511656		373.19	4.51	70.72	4 , 1	2	✓		

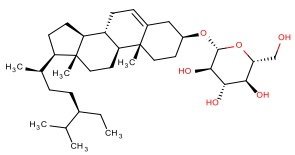
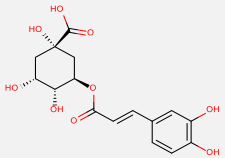
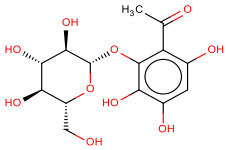
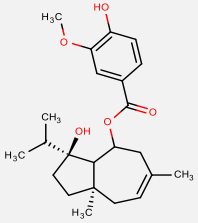
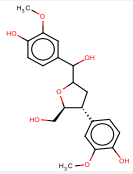
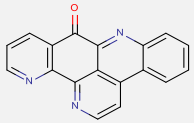
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.82, 63.36	komaroviquinone		360.44	3.27	72.83	5 , 1	2	✓		✓
-9.81, 64.44	epi-muqubillin A		392.57	6.38	55.76	4 , 1	8			✓
-9.81, 64.44	ZINC01163259		381.45	3.86	86.47	5 , 1	5	✓		✓
-9.81, 64.44	ZINC17125976		383.68	6.73	45.48	3 , 1	3			
-9.81, 64.44	ZINC17287328		533.62	3.18	116.11	6 , 2	7			✓
-9.8, 65.54	CA60		326.31	1.97	101.38	6 , 2	3	✓	✓	✓

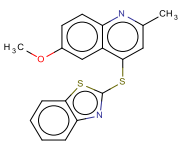
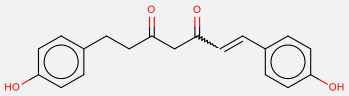
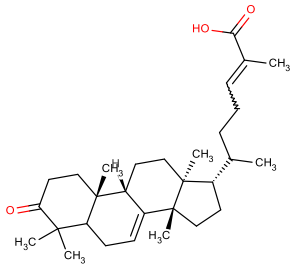
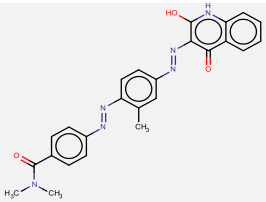
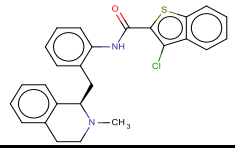
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.8, 65.54	myricetin		318.24	1.85	147.68	8 , 6	1			✓
-9.8, 65.54	6-dimethylallylaminopurine riboside		335.36	-0.43	125.55	8 , 4	5	✓		
-9.79, 66.66	ZINC17858074		359.26	5.27	35.83	3 , 1	3			
-9.78, 67.79	ZINC13154302		369.54	3.81	40.54	3 , 1	1	✓		✓
-9.77, 68.94	ZINC06511656		373.19	3.34	65.01	5 , 1	2	✓		
-9.77, 68.94	ZINC06511656		373.19	4.01	67.23	3 , 1	2	✓		

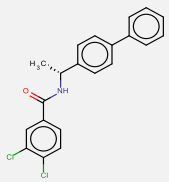
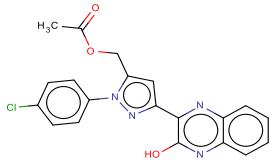
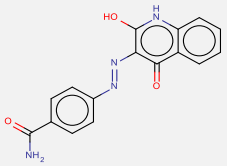
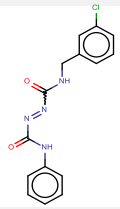
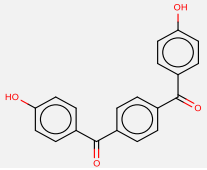
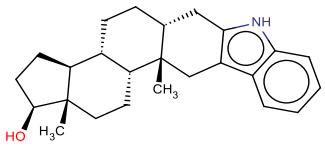
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.77, 68.94	ZINC18154478		308.29	2.86	113.65	5 , 3	3	✓		✓
-9.76, 70.12	ellagic acid		302.19	2.32	133.52	6 , 4	0	✓		
-9.76, 70.12	30-nor-lupan-3β-ol-20-one		428.69	6.33	37.3	2 , 1	1			✓
-9.76, 70.12	quercetin 7-glucosyl		464.38	-0.11	206.6	12 , 8	4		✓	✓
-9.76, 70.12	ZINC01729526		353.87	0.28	49.73	3 , 2	3	✓		
-9.75, 71.31	letestuianin A		340.37	3.84	86.99	5 , 3	7	✓		

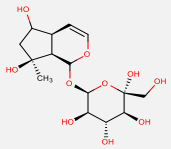
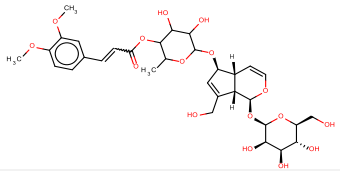
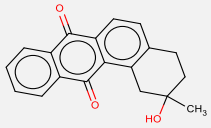
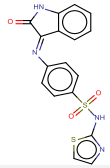
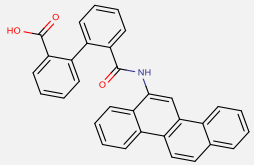
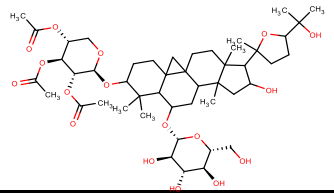
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.75, 71.31	ZINC01709784		390.44	3.53	104.16	6 , 0	8	✓		
-9.74, 72.52	ZINC01734413		341.19	5.88	43.6	3 , 0	2			
-9.74, 72.52	ZINC04726284		280.32	2.98	51.02	4 , 0	4	✓		
-9.74, 72.52	ZINC18141294		454.48	6.22	115.59	7 , 2	5			✓
-9.72, 75.01	(-)-cubebin		356.37	3.18	66.38	6 , 1	4	✓		✓
-9.72, 75.01	quercetagetin		318.24	1.85	147.68	8 , 6	1			✓

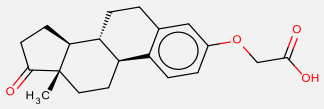
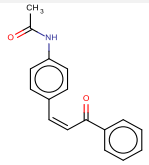
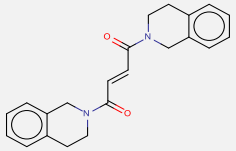
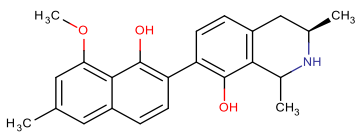
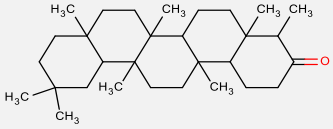
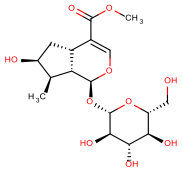
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.72, 75.01	ZINC13597346		369.54	3.81	40.54	3 , 1	1	✓		✓
-9.71, 76.29	ZINC18154478		308.29	2.97	117.14	6 , 3	3	✓		
-9.7, 77.59	violacein		343.34	1.67	94.22	3 , 4	1	✓		✓
-9.7, 77.59	ZINC17949075		402.23	5.47	75.19	4 , 2	3			✓
-9.7, 77.59	ZINC26730911		428.42	-4.61	152.45	10 , 3	3	✓	✓	
-9.69, 78.91	ZINC02476372		292.29	2.99	63.6	3 , 1	3	✓		

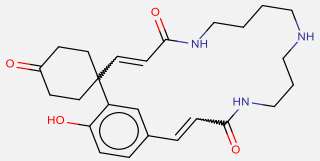
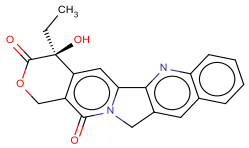
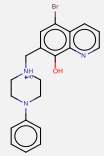
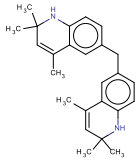
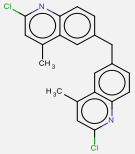
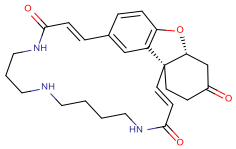
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.68, 80.25	β-sitosterol glucoside		576.85	6.07	99.38	6 , 4	9			✓
-9.67, 81.62	chlorogenic acid		354.31	-0.27	164.75	8 , 6	5			✓
-9.67, 81.62	lalioside		346.29	-1.30	177.14	10 , 7	4			
-9.67, 81.62	teferin		388.50	4.40	75.99	4 , 2	5	✓		
-9.66, 83.01	7'-hyroxy-lariciresionol		376.40	1.63	108.61	7 , 4	6	✓		✓
-9.66, 83.01	ascididemin		283.28	2.98	55.74	4 , 0	0	✓		

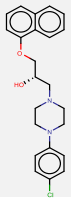
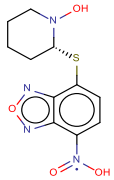
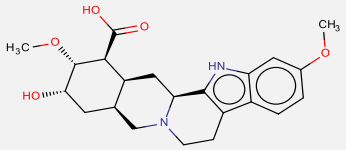
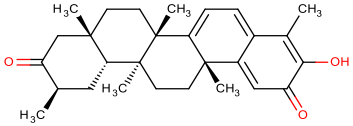
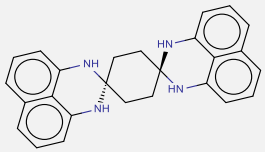
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.65, 84.42	ZINC01729578		338.45	5.12	35.01	3 , 0	3			
-9.64, 85.86	1,7-bis(4-hydroxyphenyl)-heptene-3,5-dione		310.34	4.36	74.6	4 , 2	7	✓		
-9.64, 85.86	3-oxo-tirucalla-7,24Z-dien-26-oic acid		454.68	7.51	54.37	3 , 1	5			✓
-9.64, 85.86	ZINC18141294		454.48	6.33	119.08	8 , 2	5			
-9.64, 85.86	ZINC29590277		446.99	6.73	32.34	2 , 1	4		✓	✓

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.62, 88.81	ZINC06343280		370.27	6.04	29.1	1 , 1	4			✓
-9.62, 88.81	ZINC08648354		394.81	4.29	90.13	5 , 1	5	✓		✓
-9.62, 88.81	ZINC18154478		308.29	2.97	117.14	6 , 3	3	✓		
-9.61, 90.32	CA64		316.74	2.95	82.92	4 , 2	3	✓		
-9.61, 90.32	ZINC00332365		318.32	4.28	74.6	4 , 2	4	✓		✓
-9.61, 90.32	ZINC13154305		363.54	5.21	36.02	1 , 2	0			✓

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.6, 91.86	ajugol		364.35	-3.05	169.3	10 , 7	3			✓
-9.6, 91.86	scrolepidoside		682.67	-1.19	232.52	15 , 7	12			✓
-9.58, 95.01	zenkequinone B		292.33	3.26	54.37	3 , 1	0	✓		
-9.58, 95.01	ZINC22911959		384.43	2.94	100.52	5 , 2	3	✓		✓
-9.57, 96.63	NSC128608 (CHEMBL1173811)		467.51	7.34	66.4	3 , 2	4			✓
-9.57, 96.63	acetylastrogaloside I		911.08	1.01	246.43	14 , 6	13			

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.57, 96.63	ZINC04428843		328.40	3.93	63.6	4 , 1	3	✓	✓	
-9.57, 96.63	ZINC04701260		265.31	3.13	46.17	2 , 1	4	✓		
-9.57, 96.63	ZINC04900874		346.42	2.77	40.62	2 , 0	2	✓		
-9.55, 99.94	dioncophylline B		363.45	3.51	61.72	4 , 3	2	✓		✓
-9.54, 101.65	friedelan-3-one		426.72	8.12	17.07	1 , 0	0			✓
-9.54, 101.65	8-epi-loganin		390.38	-1.84	155.14	9 , 5	5	✓		

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.54, 101.65	numismine		439.55	0.52	107.53	5 , 4	0	✓		
-9.54, 101.65	ZINC00105309		348.35	1.22	79.73	4 , 1	1	✓		
-9.54, 101.65	ZINC01707109		399.30	0.26	40.8	4 , 2	3	✓		
-9.54, 101.65	ZINC04773602		358.52	5.73	24.06	2 , 2	2			✓
-9.53, 103.38	ZINC01701460		367.27	7.06	25.78	2 , 0	2			✓
-9.52, 105.14	lunarine		437.53	1.16	96.53	5 , 3	0	✓		✓

Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.52, 105.14	ZINC29589833		396.91	4.54	35.94	4 , 1	6	✓		
-9.51, 106.92	ZINC01757986		297.31	0.29	118.9	7 , 2	3	✓		
-9.51, 106.92	ZINC12672225		400.47	-1.25	95.02	6 , 3	3	✓		
-9.5, 108.74	tingenone (maitenin)		420.58	4.92	54.37	3 , 1	0	✓		✓
-9.5, 108.74	ZINC04824645		392.50	4.49	48.12	4 , 4	0	✓		

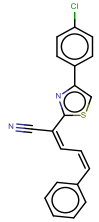
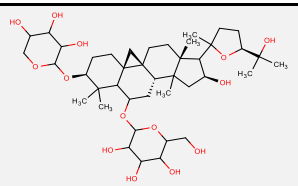
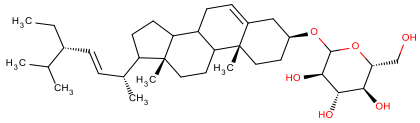
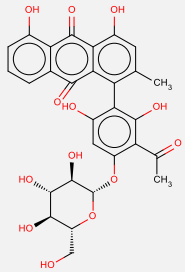
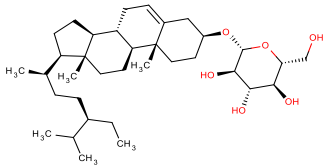
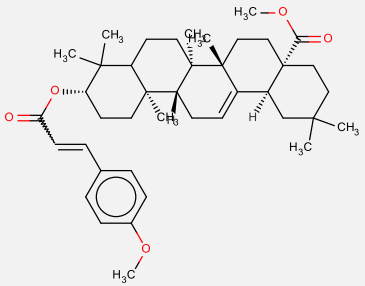
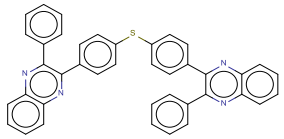
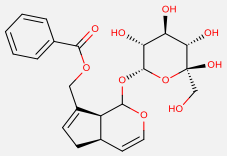
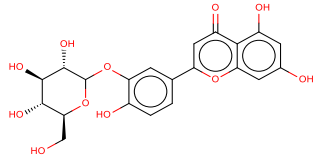
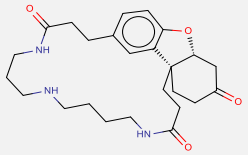
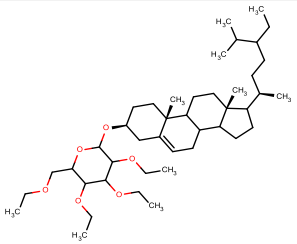
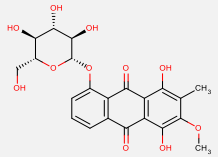
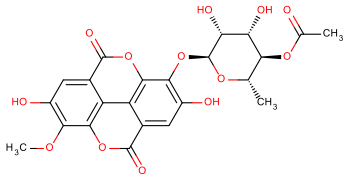
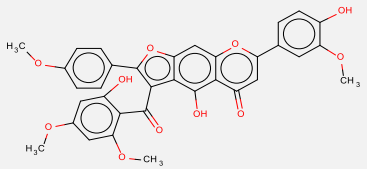
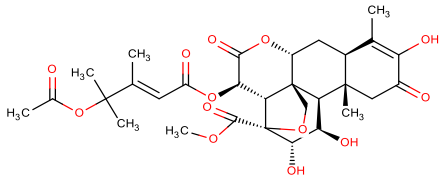
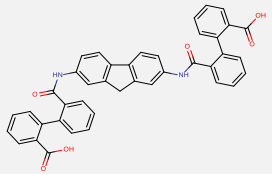
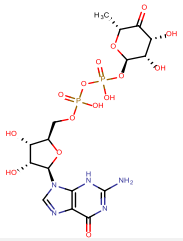
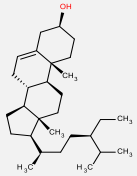
Binding energy (kcal.mol^{-1}), K_i (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1?	Hit in VS4?
-9.5, 108.74	ZINC05641037		348.85	6.23	36.68	2 , 0	4			

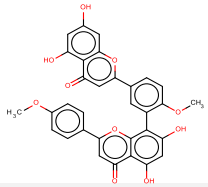
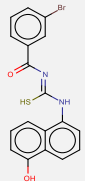
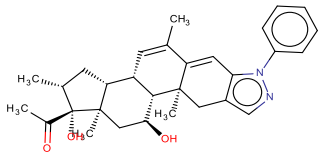
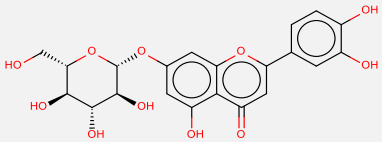
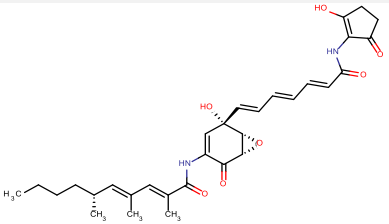
Table 7.1.9 - hits from Virtual Screen 4, listed according to predicted binding affinity.

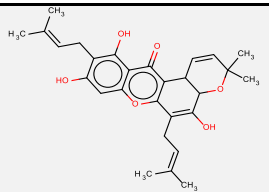
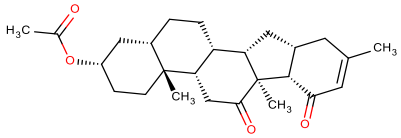
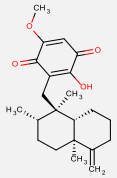
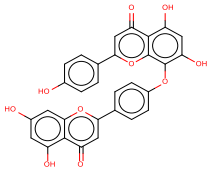
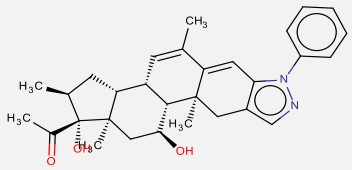
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-14.88, .01	astrogaloside III		784.97	-0.31	228.22	14 , 9	7			✓
-13.2, .21	stigmasterol-D-glucoside		574.83	5.71	99.38	6 , 4	8			✓
-12.62, .56	4'-O-demethylknipholone-4'-O-β-D-glucopyranoside		582.51	2.80	231.51	13 , 8	4			✓
-12.48, .71	β-sitosterol glucoside		576.85	6.07	99.38	6 , 4	9			✓

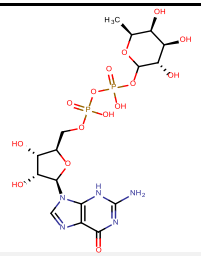
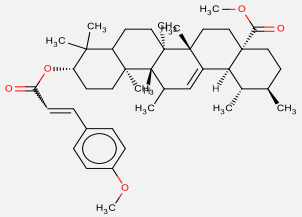
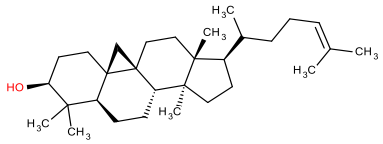
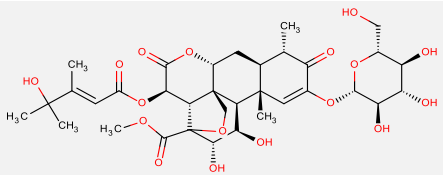
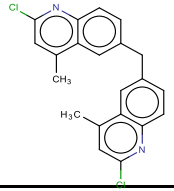
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-12.39, .83	*cis-p*-methoxycinnamoyloxyoleanolicacid methyl ester		630.90	9.62	61.83	3 , 0	7			✓
-12.3, .96	ZINC05492794		594.73	10.95	51.56	4 , 0	6			✓
-12.15, 1.24	melampyroside		450.44	0.28	155.14	9 , 5	7	✓		✓
-11.87, 1.99	luteolin-3'-O-glucoside		448.38	0.14	186.37	11 , 7	4		✓	✓
-11.83, 2.13	tetrahydrolunarine		441.56	1.08	96.53	5 , 3	0	✓		✓

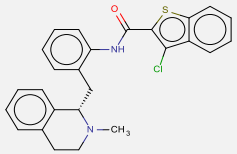
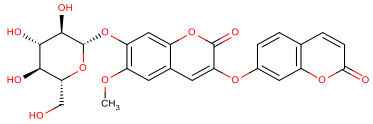
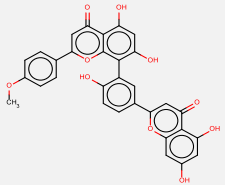
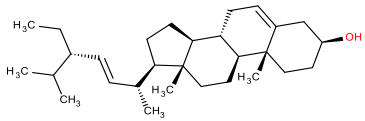
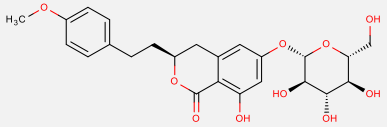
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-11.8, 2.24	tetra-acetyldaucosterol		689.06	10.07	55.38	6 , 0	17			✓
-11.76, 2.40	physcion 1-β-D-glucopyranoside		462.40	1.40	183.21	11 , 6	4			✓
-11.71, 2.61	3-O-Methylellagic acid 3'-O-α-4''-O-acetylrhamnopyranoside		504.40	0.38	187.51	10 , 4	5		✓	✓

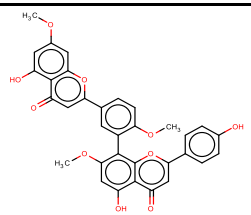
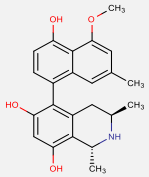
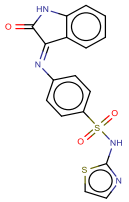
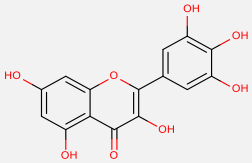
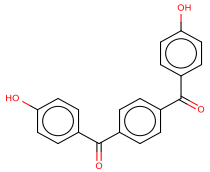
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-11.67, 2.79	cissampeloflavone		610.56	5.91	154.12	10 , 3	8			
-11.63, 2.99	bruceantinol		606.61	0.10	192.19	9 , 3	8			✓
-11.63, 2.99	NSC128590 (CHEMBL1172557)		644.67	8.53	132.8	6 , 4	8		✓	
-11.59, 3.19	GDP-4-dehydro-6-deoxy-D-mannose		587.33	-4.12	304.04	16 , 8	8		✓	✓
-11.58, 3.25	β-sitosterol		414.71	7.84	20.23	1 , 1	6			✓

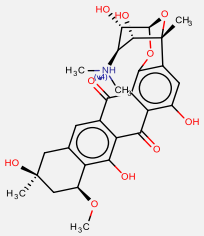
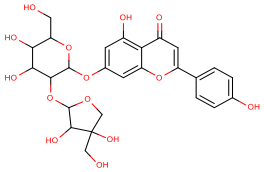
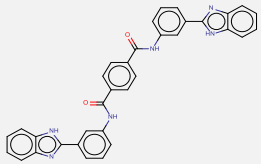
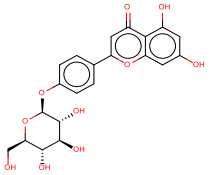
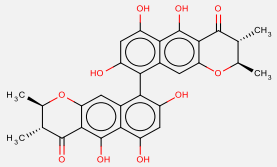
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-11.55, 3.42	isoginkgetin		566.51	5.38	151.98	10 , 4	4			✓
-11.54, 3.48	ZINC05580600		401.28	5.06	61.69	4 , 3	2			
-11.54, 3.48	ZINC17353911		472.62	4.00	75.35	4 , 2	2	✓		✓
-11.54, 3.48	luteolin-7-O-glucoside		448.38	0.14	186.37	11 , 7	4		✓	✓
-11.51, 3.66	manumycin A		550.64	3.06	145.33	7 , 4	12			✓

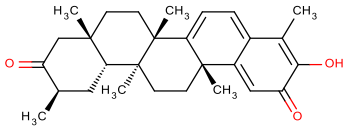
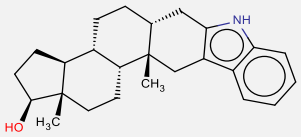
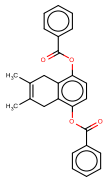
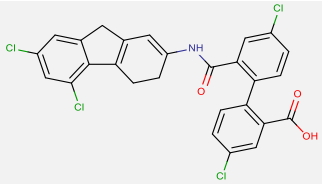
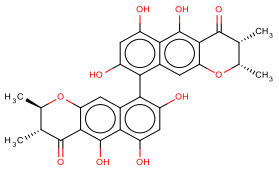
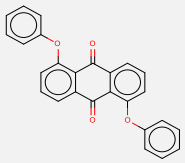
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-11.47, 3.91	tovophyllin A		464.55	5.36	96.22	6 , 3	4			✓
-11.44, 4.11	ZINC12671904		412.56	4.53	60.44	3 , 0	2	✓		✓
-11.41, 4.33	ilimaquinone		358.47	4.37	63.6	4 , 1	3	✓		✓
-11.41, 4.33	lanaroflavone		538.46	5.24	162.98	9 , 5	4			✓
-11.4, 4.40	ZINC17353912		472.62	4.00	75.35	4 , 2	2	✓		✓

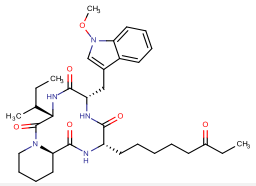
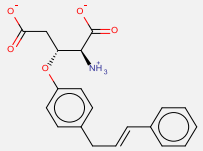
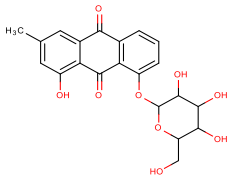
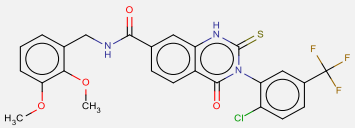
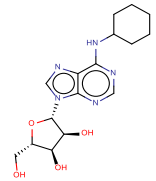
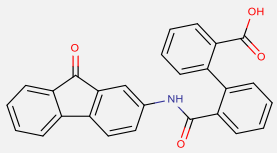
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-11.4, 4.40	GDP-L-fucose		589.34	-3.72	307.2	16 , 9	8			
-11.39, 4.48	*cis-p*-methoxycinnamoyloxyursolicacid methyl ester		644.92	9.89	61.83	3 , 0	7			✓
-11.37, 4.63	cycloartenol		426.72	7.55	20.23	1 , 1	4			✓
-11.21, 6.07	yadanzioside C		726.72	-2.35	265.27	14 , 7	9			✓
-11.2, 6.17	ZINC01701460		367.27	7.06	25.78	2 , 0	2			✓

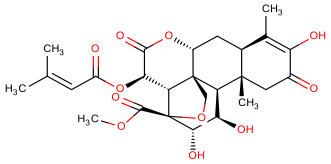
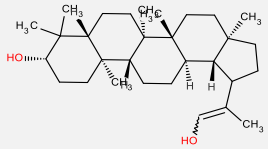
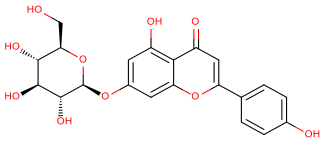
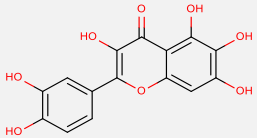
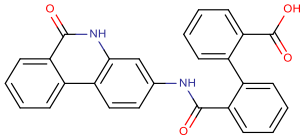
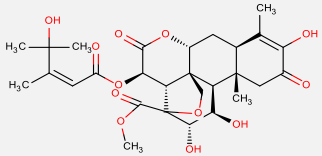
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-11.18, 6.38	ZINC29590275		446.99	6.73	32.34	2 , 1	4			
-11.18, 6.38	daphnorin		514.44	0.42	170.44	10 , 4	6			✓
-11.18, 6.38	podocarpusflavone A		552.48	5.23	162.98	10 , 5	3			✓
-11.18, 6.38	stigmasterol		412.69	7.48	20.23	1 , 1	5			✓
-11.11, 7.18	agrimonolide 6-O-β-D-glucopyranoside		476.47	1.95	155.14	9 , 5	7	✓		✓

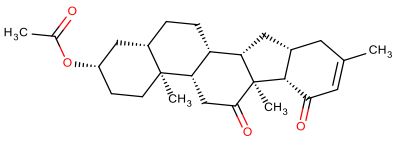
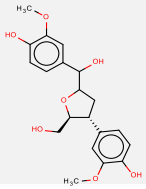
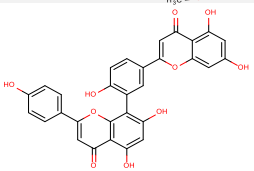
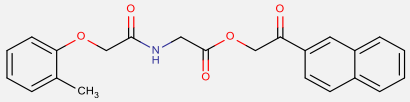
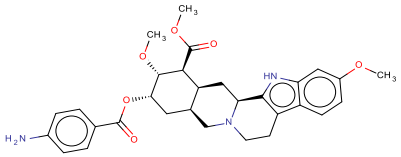
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-11.11, 7.18	ginkgetin		580.54	5.53	140.98	10 , 3	5			✓
-11.08, 7.56	korupensamine A		379.45	3.13	81.95	5 , 4	1	✓		
-11.05, 7.95	ZINC22911959		384.43	2.94	100.52	5 , 2	3	✓		✓
-11.05, 7.95	myricetin		318.24	1.85	147.68	8 , 6	1			✓
-11.04, 8.08	ZINC00332365		318.32	4.28	74.6	4 , 2	4	✓		✓

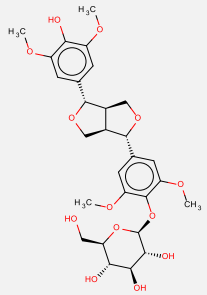
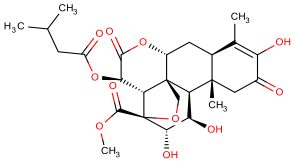
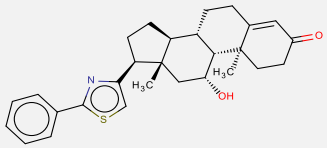
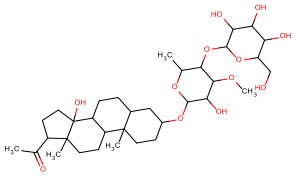
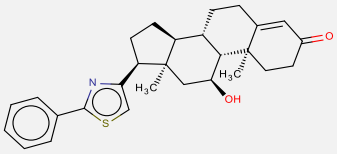
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-11.02, 8.36	ZINC11681164		542.55	-3.47	167.42	11 , 6	2			
-11.01, 8.50	apiin		564.49	-0.84	225.06	14 , 8	7			✓
-10.99, 8.79	ZINC01690699		548.59	6.78	115.56	4 , 4	6			
-10.99, 8.79	apigenin-4'-glucoside		432.38	0.44	166.14	10 , 6	4			✓
-10.98, 8.94	ZINC17465983		546.52	5.67	173.98	10 , 6	0			✓

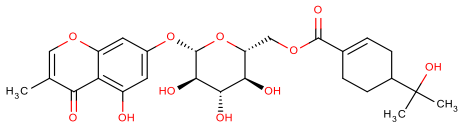
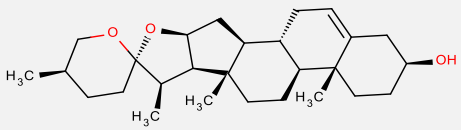
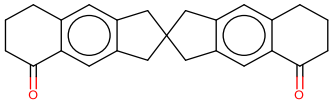
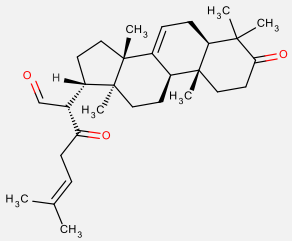
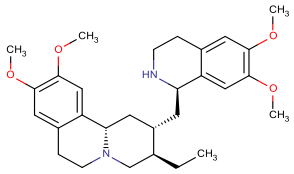
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.97, 9.10	tingenone (maitenin)		420.58	4.92	54.37	3 , 1	0	✓		✓
-10.94, 9.57	ZINC13154305		363.54	5.21	36.02	1 , 2	0			✓
-10.92, 9.90	ZINC01755448		398.45	6.87	52.6	2 , 0	6		✓	✓
-10.9, 10.24	CHEMBL1172759		545.24	7.24	66.4	3 , 2	4			✓
-10.89, 10.41	ZINC05462670		546.52	5.67	173.98	10 , 6	0			
-10.88, 10.59	ZINC03875548		392.40	5.92	52.6	2 , 0	4		✓	✓

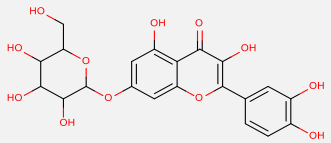
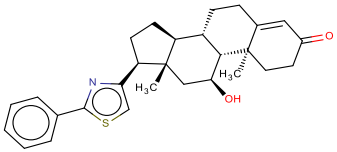
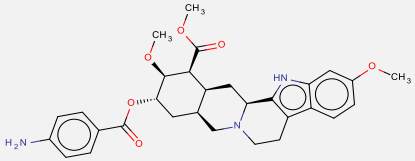
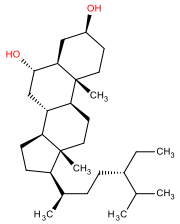
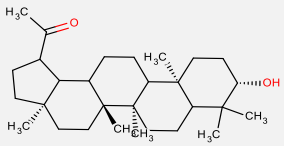
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.82, 11.72	apicidin		623.78	3.37	138.84	6 , 3	12			
-10.81, 11.92	ZINC00600877		354.38	0.87	117.13	5 , 1	9	✓	✓	
-10.81, 11.92	chrysophanol-8-glucoside		416.38	1.21	153.75	9 , 5	3	✓		✓
-10.81, 11.92	Qc2		549.95	5.48	79.9	4 , 2	6			✓
-10.8, 12.12	ZINC00001761		349.38	0.01	125.55	8 , 4	4	✓	✓	✓
-10.8, 12.12	NSC67692 (CHEMBL1172978)		419.43	5.50	83.47	4 , 2	4			✓

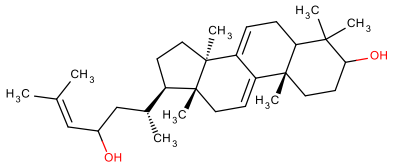
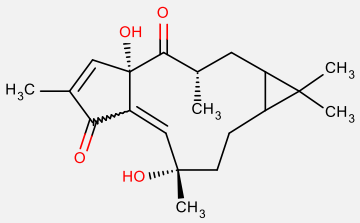
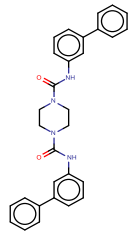
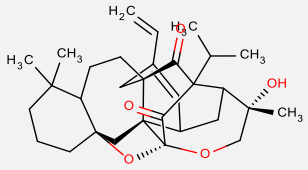
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.79, 12.33	brusatol		520.53	0.24	165.89	8 , 3	5		✓	✓
-10.78, 12.54	hennadiol		442.72	6.86	40.46	2 , 2	1			✓
-10.77, 12.75	apigenin-7-glucoside		432.38	0.44	166.14	10 , 6	4		✓	✓
-10.76, 12.97	quercetagetin		318.24	1.85	147.68	8 , 6	1			✓
-10.76, 12.97	NSC127133 (CHEMBL1172911)		434.44	5.14	95.5	4 , 3	4			✓
-10.73, 13.64	bruceine C		564.58	-0.35	186.12	9 , 4	6			✓

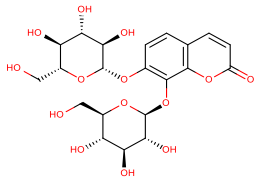
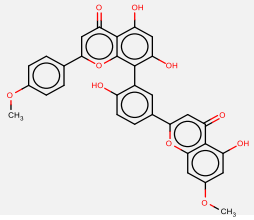
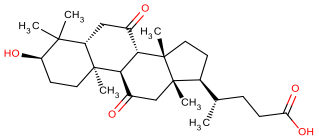
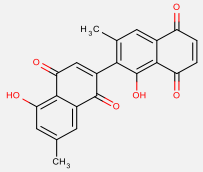
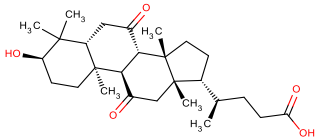
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.69, 14.59	ZINC12671893		412.56	4.53	60.44	3 , 0	2	✓		✓
-10.65, 15.61	7'-hyroxy-lariciresionol		376.40	1.63	108.61	7 , 4	6	✓		✓
-10.65, 15.61	amentoflavone		538.46	5.09	173.98	10 , 6	2			✓
-10.64, 15.88	CHEMBL1172971		391.42	3.07	81.7	4 , 1	9	✓		
-10.63, 16.15	ZINC17287328		533.62	3.18	116.11	6 , 2	7			✓

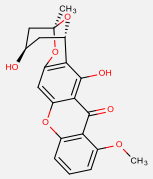
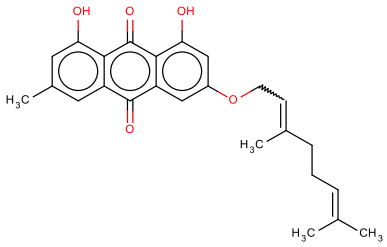
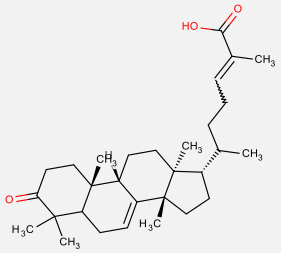
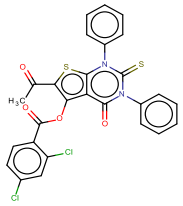
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.63, 16.15	(+)-Syringaresinol O-β-D-glucopyranoside		580.58	-0.31	174.99	13 , 5	9			
-10.61, 16.70	bruceine A		522.54	0.05	165.89	8 , 3	6			✓
-10.6, 16.99	ZINC13099048		447.63	5.52	50.19	3 , 1	2			✓
-10.59, 17.28	caratuberside A		656.80	1.05	184.6	12 , 6	7			✓
-10.58, 17.57	ZINC13099050		447.63	5.52	50.19	3 , 1	2			✓

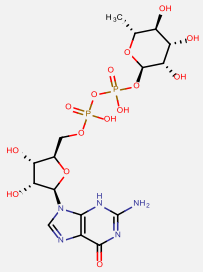
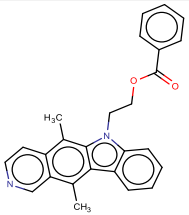
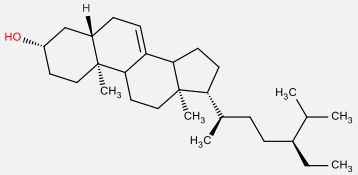
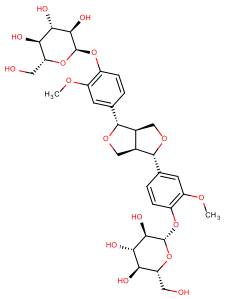
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.58, 17.57	camaldulenside		520.53	2.10	172.21	10 , 5	7		✓	✓
-10.57, 17.87	nitogenin (diosgenin)		414.62	4.93	38.69	3 , 1	0	✓		✓
-10.55, 18.48	ZINC01855333		356.46	5.37	34.14	2 , 0	0			✓
-10.53, 19.12	3,23-dioxotirucalla-7,24-dien-21-al		452.67	6.42	51.21	3 , 0	5			✓
-10.53, 19.12	emetine		480.64	4.49	52.19	6 , 1	7	✓		✓

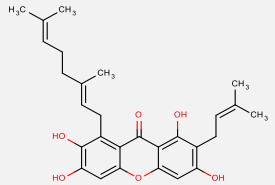
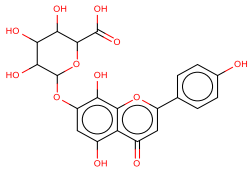
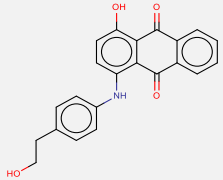
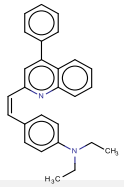
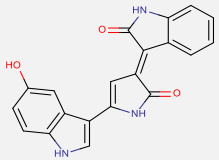
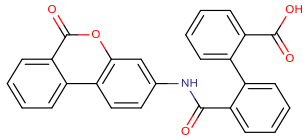
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.52, 19.44	quercetin 7-glucosyl		464.38	-0.11	206.6	12 , 8	4		✓	✓
-10.5, 20.11	ZINC13099051		447.63	5.52	50.19	3 , 1	2			✓
-10.49, 20.45	ZINC17287324		533.62	3.18	116.11	6 , 2	7			
-10.48, 20.80	5alpha-poriferastane-3β,6α-diol		432.72	6.94	40.46	2 , 2	6			✓
-10.47, 21.15	30-nor-lupan-3β-ol-20-one		428.69	6.33	37.3	2 , 1	1			✓

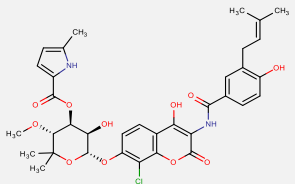
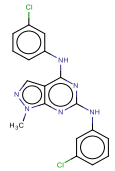
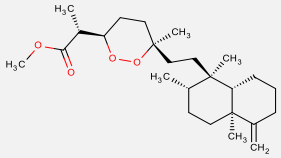
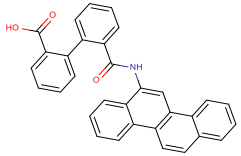
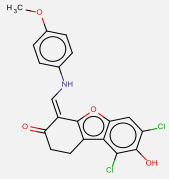
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.46, 21.51	5α-Lanosta-7,9(11),24-triene-3α,23-diol		440.70	6.11	40.46	2 , 2	4			
-10.45, 21.88	15-epi-4Z-jatrogrossidentadion		332.43	3.10	74.6	4 , 2	0	✓		
-10.43, 22.63	ZINC04783229		476.57	5.60	64.68	2 , 2	4			✓
-10.43, 22.63	hydrangdione A		494.66	5.98	72.83	5 , 1	2			

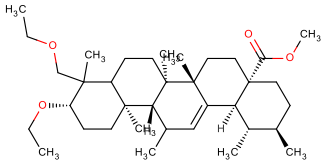
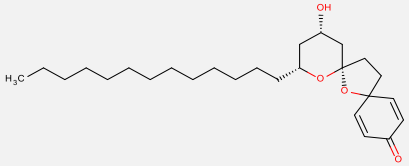
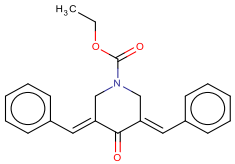
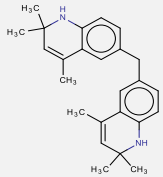
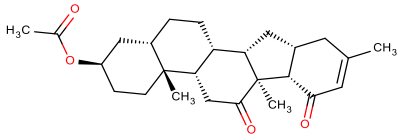
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.43, 22.63	daphneside		502.42	-3.36	225.06	13 , 8	6		✓	
-10.43, 22.63	podocarpusflavone B		566.51	5.38	151.98	10 , 4	4			✓
-10.41, 23.41	ZINC12670903		446.62	4.22	91.67	5 , 2	4	✓		
-10.4, 23.81	diospyrin		374.34	4.40	108.74	6 , 2	1	✓		
-10.37, 25.04	ZINC12670920		446.62	4.22	91.67	5 , 2	4	✓		

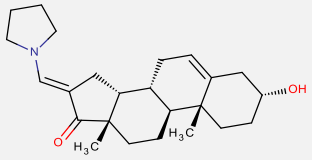
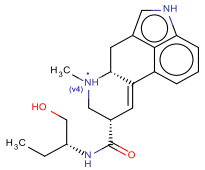
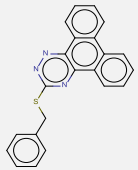
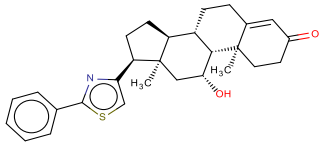
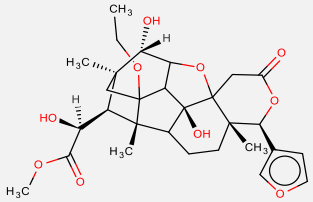
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.37, 25.04	chaetoxanthone A		370.35	2.86	94.45	6 , 2	1	✓		
-10.37, 25.04	3-geranyloxy-6-methyl-1,8-dihydroxyanthraquinone		406.47	6.99	83.83	5 , 2	6			✓
-10.37, 25.04	3-oxo-tirucalla-7,24Z-dien-26-oic acid		454.68	7.51	54.37	3 , 1	5			✓
-10.36, 25.47	ZINC01644254		567.46	7.55	66.92	3 , 0	6			✓

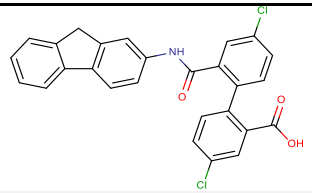
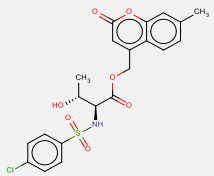
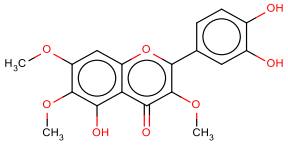
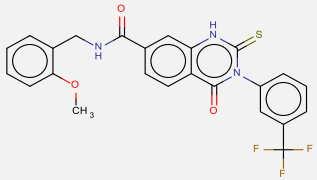
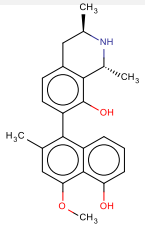
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.36, 25.47	GDP-rhamnose		589.34	-3.72	307.2	16 , 9	8			
-10.35, 25.90	ZINC01635676		394.47	5.92	44.12	2 , 0	5			
-10.35, 25.90	24-ethyl-cholest-5a-7-en-3-β-ol		414.71	7.84	20.23	1 , 1	6			
-10.35, 25.90	pinoresinol di-O-β-D-glucopyranoside		682.67	-2.26	235.68	16 , 8	10			✓

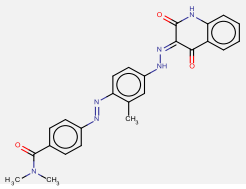
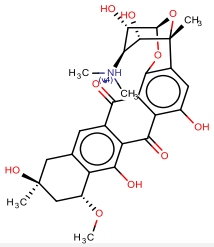
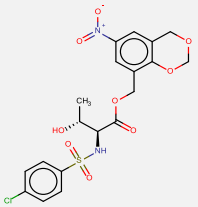
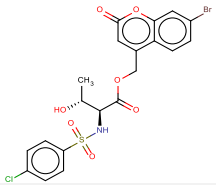
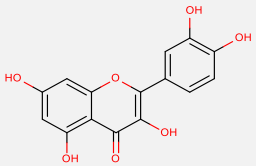
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.35, 25.90	norcowanin		464.55	7.51	107.22	5 , 4	7			
-10.34, 26.34	isoscutellarin		462.36	0.46	203.44	12 , 7	4		✓	✓
-10.32, 27.25	ZINC05124960		359.37	5.53	86.63	5 , 3	4			✓
-10.32, 27.25	ZINC08581460		378.51	7.17	16.13	2 , 0	6			✓
-10.32, 27.25	violacein		343.34	1.67	94.22	3 , 4	1	✓		✓
-10.32, 27.25	NSC128604 (CHEMBL1173036)		435.43	5.32	92.7	4 , 2	4			✓

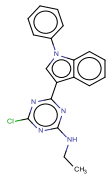
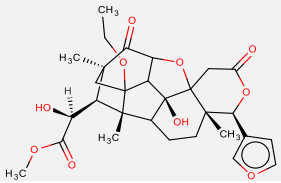
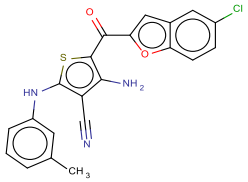
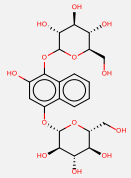
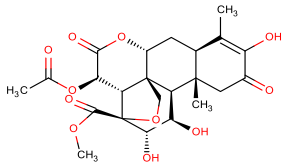
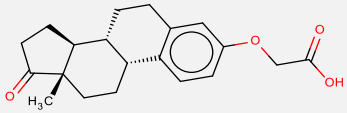
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.32, 27.25	clorobiocin		697.13	4.94	185.87	9 , 5	10			
-10.31, 27.71	ZINC01718486		385.25	5.23	67.66	5 , 2	4		✓	
-10.31, 27.71	sigmosceptrellin A methyl ester		406.60	6.27	44.76	3 , 0	6			✓
-10.31, 27.71	NSC128608 (CHEMBL1173811)		467.51	7.34	66.4	3 , 2	4			✓
-10.3, 28.18	ZINC05487838		404.24	4.51	71.7	4 , 2	3	✓		

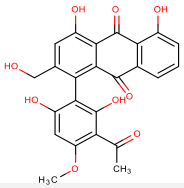
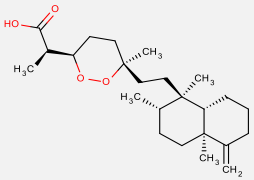
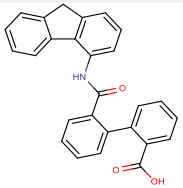
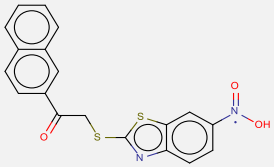
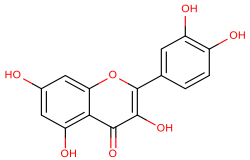
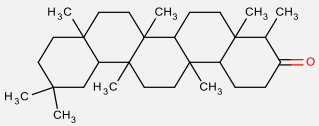
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.3, 28.18	methyl 3β,23-diacetoxy-12-ursen-28-oate		556.86	7.73	44.76	3 , 0	7			
-10.29, 28.66	aculeatin D		418.61	7.02	55.76	4 , 1	12			✓
-10.28, 29.15	ZINC05709639		347.41	4.49	46.61	2 , 0	4	✓		
-10.27, 29.65	ZINC04773602		358.52	5.73	24.06	2 , 2	2			✓
-10.25, 30.67	ZINC12671898		412.56	4.53	60.44	3 , 0	2	✓		✓

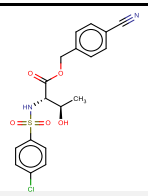
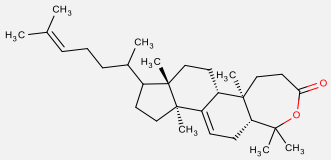
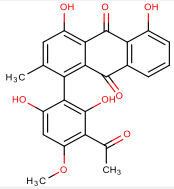
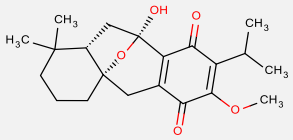
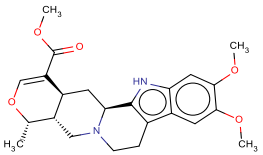
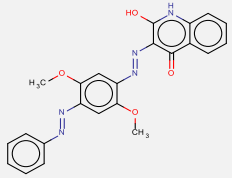
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.24, 31.19	ZINC13154302		369.54	3.81	40.54	3 , 1	1	✓		✓
-10.23, 31.72	ZINC01081577		340.44	-2.52	69.56	3 , 4	4	✓		
-10.23, 31.72	ZINC01573829		353.44	5.74	38.67	3 , 0	3		✓	✓
-10.23, 31.72	ZINC13099047		447.63	5.52	50.19	3 , 1	2			
-10.23, 31.72	1-O-deacetylkhayanolide B		546.61	0.71	144.89	7 , 3	6			

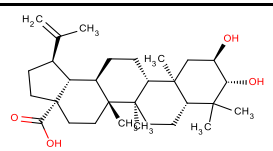
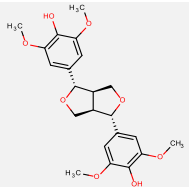
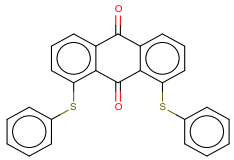
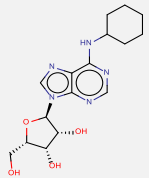
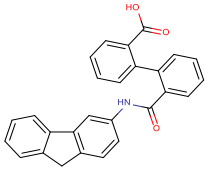
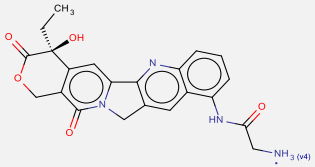
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.23, 31.72	NSC132249 (CHEMBL1173812)		474.34	7.34	66.4	3 , 2	4			✓
-10.22, 32.26	ZINC03313499		465.90	2.80	119	5 , 2	7	✓		
-10.22, 32.26	chrysosplenol D		360.31	2.26	114.68	8 , 3	4	✓		
-10.21, 32.81	Qc3		485.48	5.04	70.67	3 , 2	6			
-10.2, 33.37	5'-O-demethyl-ent-dioncophylleine A		363.45	3.48	61.72	4 , 3	1	✓		

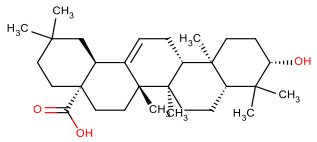
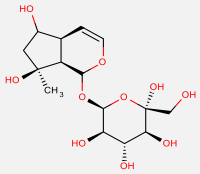
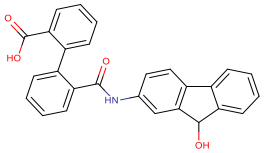
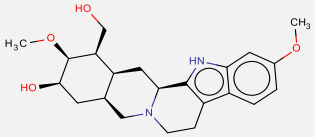
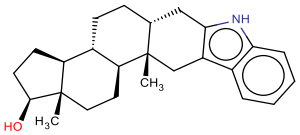
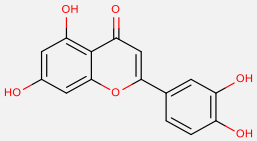
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.19, 33.93	ZINC18141294		454.48	6.22	115.59	7 , 2	5			✓
-10.18, 34.51	ZINC11681161		542.55	-3.47	167.42	11 , 6	2			
-10.18, 34.51	ZINC03314942		486.88	2.29	154.3	8 , 2	8	✓		
-10.18, 34.51	ZINC03417739		530.77	3.06	119	5 , 2	7			✓
-10.18, 34.51	quercetin		302.24	2.16	127.45	7 , 5	1	✓	✓	✓

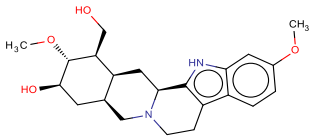
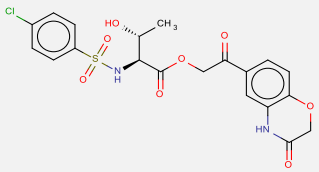
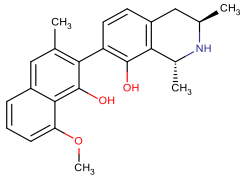
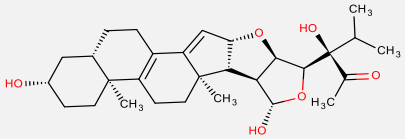
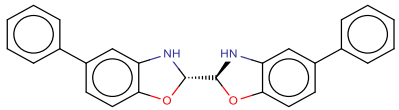
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.17, 35.10	ZINC00161700		349.82	5.55	55.63	4 , 1	4			
-10.17, 35.10	khayanolide E		544.59	1.47	141.73	7 , 2	6			
-10.16, 35.70	ZINC01568966		407.87	5.83	92.05	4 , 2	4		✓	✓
-10.16, 35.70	lawsoniaside		500.45	-2.48	218.99	13 , 9	6		✓	✓
-10.16, 35.70	bruceine B		480.46	-1.38	165.89	8 , 3	4	✓		
-10.15, 36.30	ZINC13099025		328.40	3.93	63.6	4 , 1	3	✓		✓

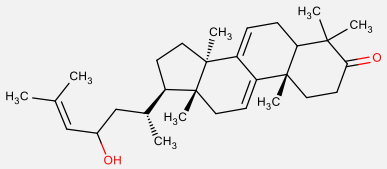
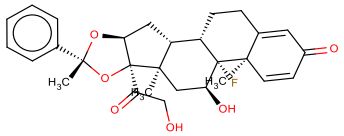
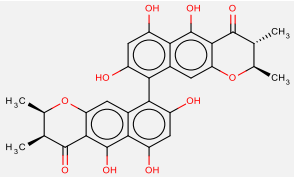
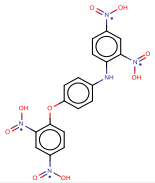
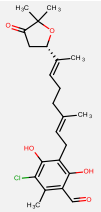
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.15, 36.30	gaboroquinone B		450.39	3.93	161.59	9 , 5	3	✓		
-10.15, 36.30	sigmosceptrellin A		392.57	6.12	55.76	4 , 1	5			✓
-10.15, 36.30	NSC128596 (CHEMBL1173452)		405.44	6.14	66.4	3 , 2	4			
-10.14, 36.92	ZINC01163259		381.45	3.86	86.47	5 , 1	5	✓		✓
-10.14, 36.92	quercetin*		302.24	2.16	127.45	7 , 5	1	✓		
-10.13, 37.55	friedelan-3-one		426.72	8.12	17.07	1 , 0	0			✓

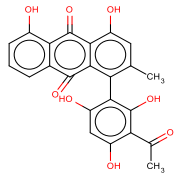
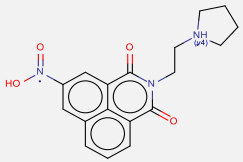
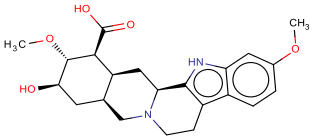
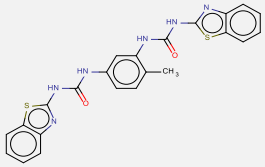
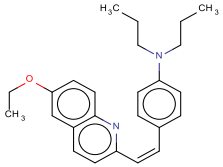
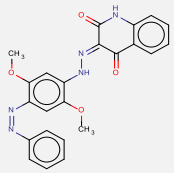
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.12, 38.19	ZINC04286641		408.86	2.55	116.49	5 , 2	7	✓		
-10.11, 38.84	etandrolide		440.70	7.54	26.3	1 , 0	4			
-10.1, 39.50	knipholone		434.39	5.22	141.36	8 , 4	2			✓
-10.1, 39.50	komaroviquinone		360.44	3.27	72.83	5 , 1	2	✓		✓
-10.09, 40.17	ZINC04777937		412.48	2.24	73.02	5 , 1	4	✓		
-10.09, 40.17	ZINC18169477		429.43	6.21	117.23	9 , 2	6			

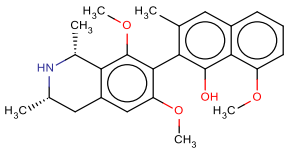
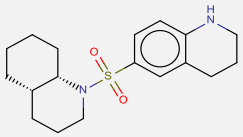
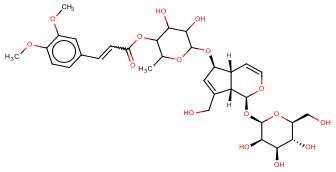
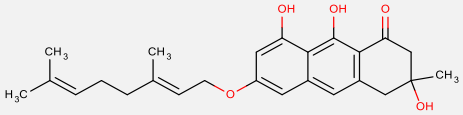
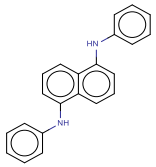
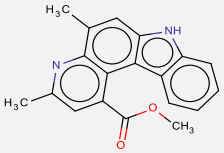
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.09, 40.17	2α-hydroxybetulinic acid		472.70	5.56	77.76	4 , 3	2			
-10.09, 40.17	(+)-syringaresinol		418.44	1.96	95.84	8 , 2	6	✓		✓
-10.08, 40.86	ZINC05124931		424.53	7.30	34.14	2 , 0	4			✓
-10.08, 40.86	ZINC00057276		349.38	0.01	125.55	8 , 4	4	✓		
-10.08, 40.86	NSC128594 (CHEMBL1173451)		405.44	6.14	66.4	3 , 2	4			✓
-10.07, 41.55	ZINC01625746		421.43	-5.26	136.47	6 , 3	3	✓		

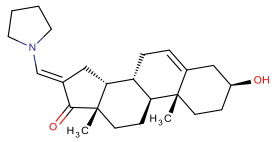
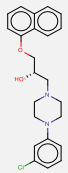
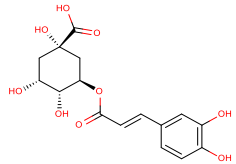
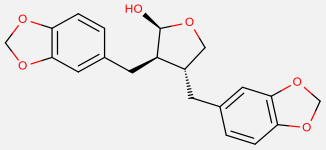
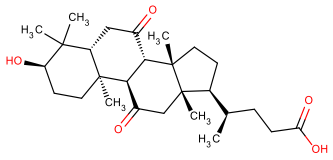
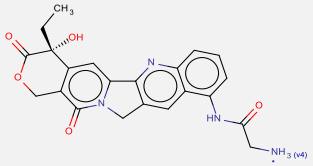
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.06, 42.26	oleanolic acid		456.70	6.59	57.53	3 , 2	1			
-10.04, 43.71	ajugol		364.35	-3.05	169.3	10 , 7	3			✓
-10.04, 43.71	NSC127134 (CHEMBL1172616)		421.44	5.06	86.63	4 , 3	4			
-10.03, 44.45	ZINC13145029		386.48	0.95	77.95	5 , 3	3	✓		
-10.03, 44.45	ZINC13154307		363.54	5.21	36.02	1 , 2	0			✓
-10.03, 44.45	luteolin		286.24	2.40	107.22	6 , 4	1	✓		

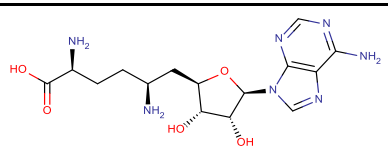
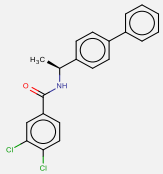
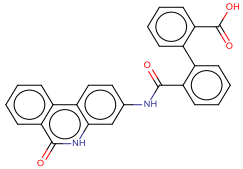
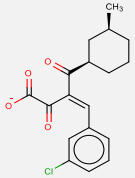
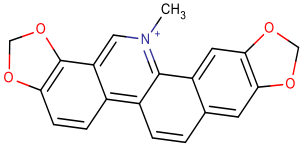
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.02, 45.21	ZINC13145030		386.48	0.95	77.95	5 , 3	3	✓		
-10.02, 45.21	ZINC03450331		482.89	0.98	148.1	7 , 3	8	✓		
-10.02, 45.21	dioncophylline E		363.45	3.51	61.72	4 , 3	1	✓		
-10.02, 45.21	vernoguinosterol		486.64	2.84	96.22	6 , 3	3	✓		✓
-10.01, 45.98	ZINC01736227		392.45	6.01	42.52	4 , 2	3			✓

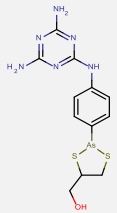
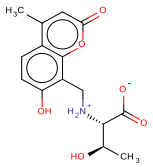
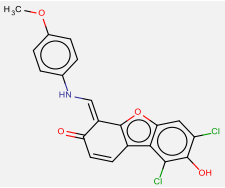
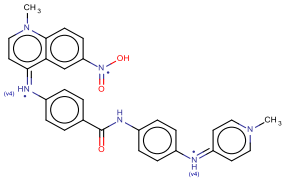
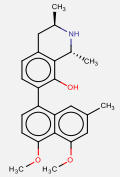
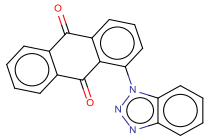
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-10.01, 45.98	23-Hydroxy-5α-lanosta-7,9(11),24-triene-3-one		438.69	6.68	37.3	2 , 1	4			✓
-10, 46.76	ZINC11677172		496.57	3.65	93.06	6 , 2	3	✓		
-10, 46.76	ZINC17465979		546.52	5.67	173.98	10 , 6	0			✓
-9.99, 47.56	ZINC17465965		445.34	0.31	247.3	13 , 5	8			
-9.99, 47.56	ascofuranone		420.93	6.40	83.83	5 , 2	7			✓

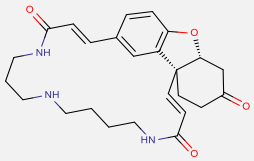
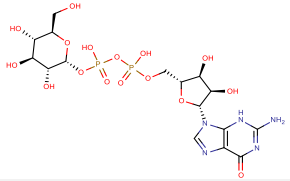
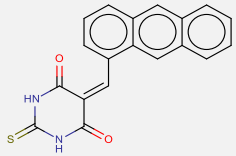
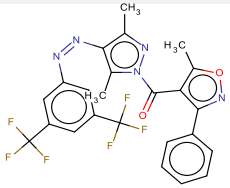
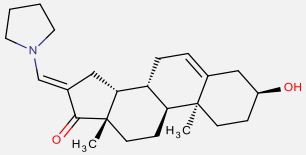
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.99, 47.56	4'-O-demethylknipholone		420.37	5.72	152.36	8 , 5	1			✓
-9.98, 48.37	ZINC04217305		341.36	-3.47	98.33	6 , 2	4	✓		✓
-9.98, 48.37	ZINC12672231		400.47	-1.25	95.02	6 , 3	3	✓		
-9.97, 49.19	ZINC01572309		474.56	6.43	108.04	4 , 4	4			✓
-9.97, 49.19	ZINC05105895		374.52	6.77	25.36	3 , 0	9			
-9.95, 50.88	ZINC18169477		429.43	6.10	113.74	8 , 2	6			

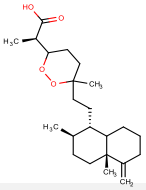
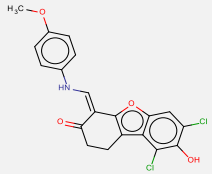
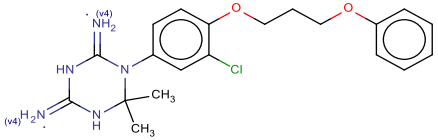
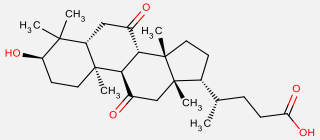
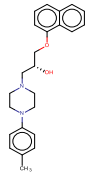
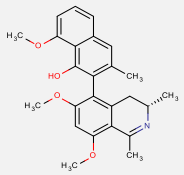
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.95, 50.88	ancistrotectonine		407.50	4.08	59.95	5 , 2	3	✓		
-9.94, 51.75	ZINC05493736		334.48	3.12	49.41	3 , 1	1	✓	✓	✓
-9.93, 52.63	scrolepidoside		682.67	-1.19	232.52	15 , 7	12			✓
-9.93, 52.63	vismione D		410.50	5.23	86.99	5 , 3	6			✓
-9.92, 53.52	ZINC01608817		310.39	5.84	24.06	2 , 2	4			
-9.92, 53.52	ZINC01691782		304.34	3.90	54.98	2 , 1	2	✓		

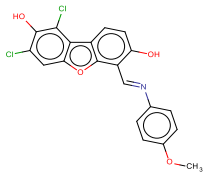
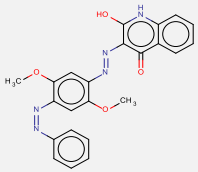
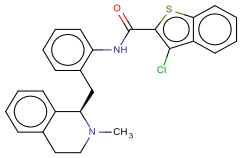
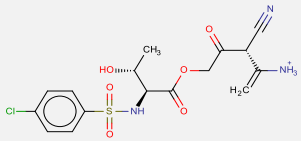
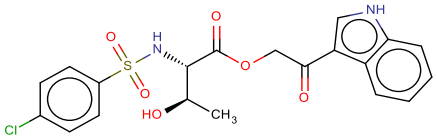
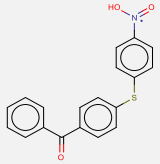
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.92, 53.52	ZINC13597348		369.54	3.81	40.54	3 , 1	1	✓		✓
-9.92, 53.52	ZINC23118772		396.91	4.54	35.94	4 , 1	6	✓		
-9.92, 53.52	chlorogenic acid		354.31	-0.27	164.75	8 , 6	5			✓
-9.92, 53.52	(-)-cubebin		356.37	3.18	66.38	6 , 1	4	✓		✓
-9.91, 54.43	ZINC12670909		446.62	4.22	91.67	5 , 2	4	✓		
-9.91, 54.43	ZINC13000556		421.43	-5.26	136.47	6 , 3	3	✓		✓

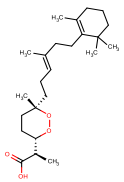
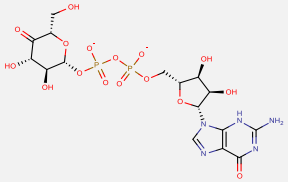
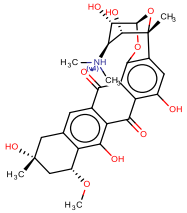
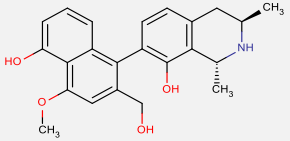
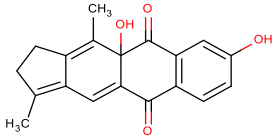
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.91, 54.43	sinefungin		381.39	-5.14	208.65	11 , 6	7			✓
-9.9, 55.36	ZINC01735767		370.27	6.04	29.1	1 , 1	4			✓
-9.89, 56.30	ZINC29589888		434.44	5.14	95.5	4 , 3	4		✓	
-9.89, 56.30	ZINC01624619		333.79	5.35	74.27	4 , 0	5			
-9.89, 56.30	sanguinarine		332.33	-0.94	40.8	4 , 0	0	✓		

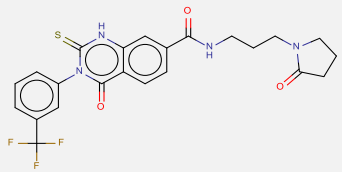
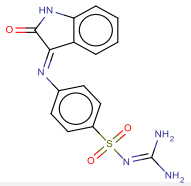
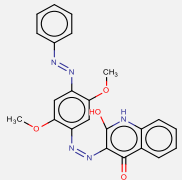
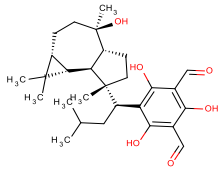
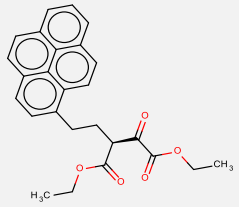
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.89, 56.30	melarsoprol		398.34	1.34	122.97	7 , 4	4	✓		
-9.86, 59.23	ZINC01856861		307.30	-1.72	123.5	5 , 3	5	✓		
-9.85, 60.24	ZINC17949075		402.23	4.59	71.7	4 , 2	3	✓		
-9.85, 60.24	ZINC18057104		507.56	3.89	130.51	8 , 4	5			
-9.85, 60.24	dioncophylline D		377.48	3.80	50.72	4 , 2	3	✓		
-9.84, 61.26	ZINC04896472		325.32	4.03	64.85	4 , 0	1	✓	✓	✓

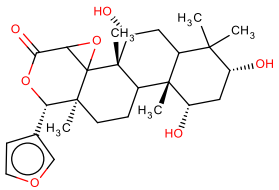
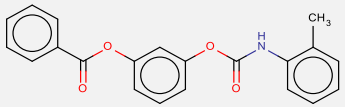
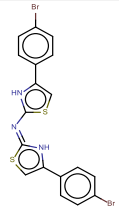
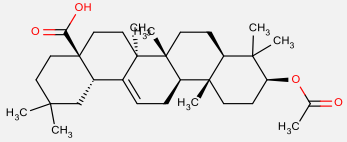
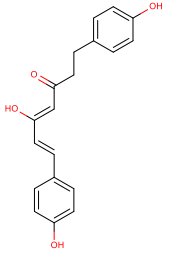
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.84, 61.26	lunarine		437.53	1.16	96.53	5 , 3	0	✓		✓
-9.84, 61.26	GDP-L-glucose		605.34	-4.77	327.43	17 , 10	9			✓
-9.83, 62.30	ZINC01580992		332.38	3.65	58.2	2 , 2	1	✓		
-9.83, 62.30	ZINC05664697		521.41	6.54	85.64	5 , 0	6			
-9.83, 62.30	ZINC13597346		369.54	3.81	40.54	3 , 1	1	✓		✓

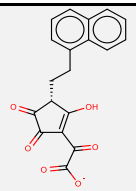
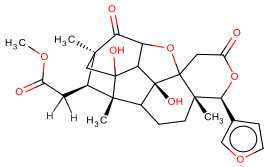
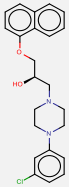
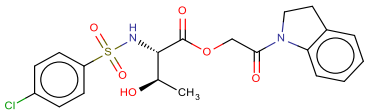
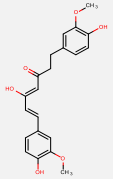
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.83, 62.30	sigmosceptrellin B		378.55	5.82	55.76	4 , 1	5			✓
-9.82, 63.36	ZINC05487838		404.24	4.51	71.7	4 , 2	3	✓	✓	✓
-9.82, 63.36	ZINC16951318		403.91	4.59	84.18	7 , 4	7	✓		
-9.81, 64.44	ZINC12670914		446.62	4.22	91.67	5 , 2	4	✓		✓
-9.81, 64.44	ZINC29589835		376.49	4.45	35.94	4 , 1	6	✓		
-9.81, 64.44	ancistrotanzanine A		405.49	4.60	60.28	5 , 1	3	✓		

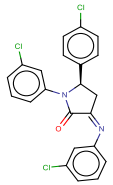
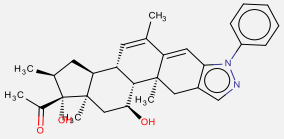
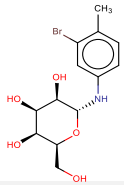
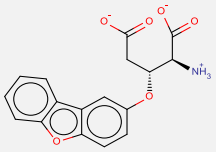
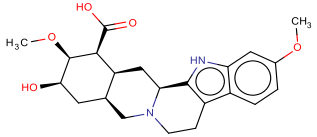
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.8, 65.54	ZINC17949075		402.23	5.47	75.19	4 , 2	3		✓	✓
-9.8, 65.54	ZINC18169477		429.43	6.21	117.23	9 , 2	6			
-9.8, 65.54	ZINC29590277		446.99	6.73	32.34	2 , 1	4		✓	✓
-9.8, 65.54	ZINC02629278		416.86	0.10	161.2	6 , 3	9	✓		
-9.8, 65.54	ZINC03315231		450.89	2.30	125.56	5 , 3	8	✓		
-9.78, 67.79	ZINC01675858		336.38	4.15	73.58	4 , 1	5	✓		

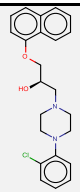
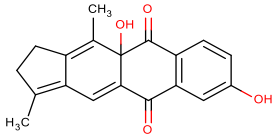
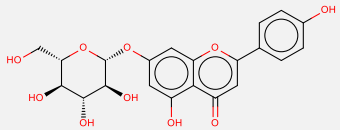
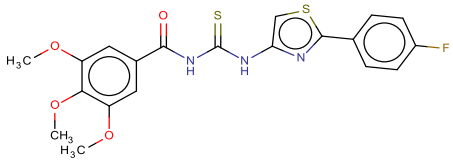
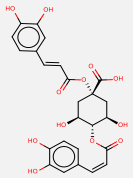
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.78, 67.79	epi-muqubillin A		392.57	6.38	55.76	4 , 1	8			✓
-9.78, 67.79	GDP-β-L-keto-galactose		601.31	-5.17	329.93	17 , 7	9			
-9.77, 68.94	ZINC11681166		542.55	-3.47	167.42	11 , 6	2			✓
-9.76, 70.12	dioncopeltine A		379.45	2.19	81.95	5 , 4	2	✓		
-9.76, 70.12	kigelinol		308.33	1.83	74.6	4 , 2	0	✓		✓

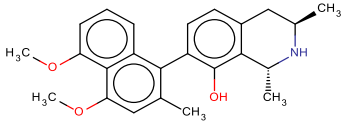
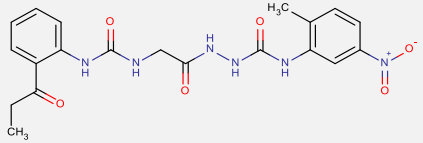
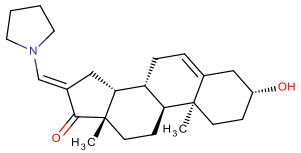
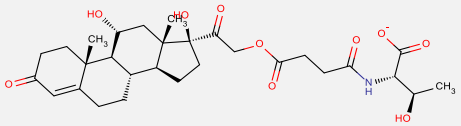
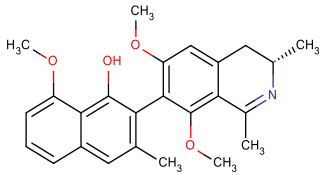
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.76, 70.12	Qc5		490.50	3.00	81.75	3 , 2	7	✓		
-9.75, 71.31	ZINC16951320		343.36	1.05	140	7 , 3	2	✓		
-9.74, 72.52	ZINC18169477		429.43	6.21	117.23	9 , 2	6			✓
-9.74, 72.52	macrocarpal B		472.61	7.63	115.06	6 , 4	6			
-9.73, 73.76	ZINC01593340		416.47	6.07	69.67	3 , 0	10			

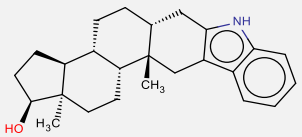
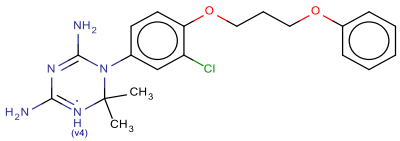
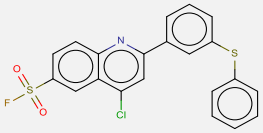
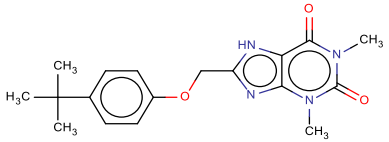
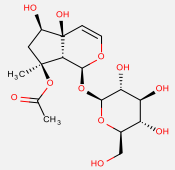
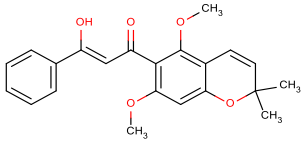
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.73, 73.76	1α,3α,7α-trideacetylkhivorin		460.56	1.99	112.66	5 , 3	1	✓		✓
-9.71, 76.29	ZINC00398390		347.36	5.66	64.63	3 , 1	6			
-9.71, 76.29	ZINC13154298		494.25	6.82	40.18	2 , 2	3			
-9.71, 76.29	acetyloleanolic acid		498.74	7.04	63.6	3 , 1	3			
-9.71, 76.29	(4Z,6E)-5-hydroxy-1,7-bis(4-hydroxyphenyl)hepta-4,6-dien-3-one		310.34	3.99	77.76	4 , 3	6	✓		✓

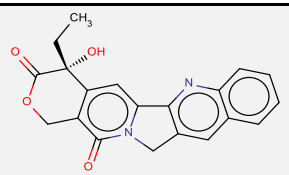
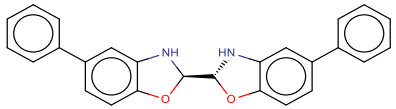
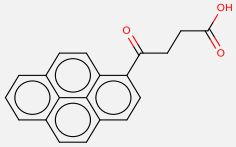
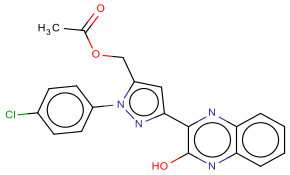
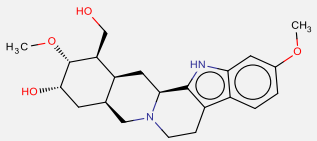
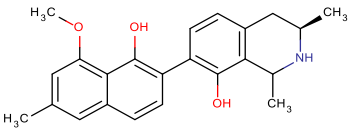
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.7, 77.59	ZINC01610563		337.30	3.65	111.57	6 , 1	5	✓		
-9.7, 77.59	6-dehydroxykhayanolide E		500.54	1.27	132.5	6 , 2	4			✓
-9.69, 78.91	ZINC23118770		396.91	4.54	35.94	4 , 1	6	✓		
-9.69, 78.91	ZINC03365068		452.91	1.79	113.01	5 , 2	7	✓		
-9.68, 80.25	letestuianin B		370.40	3.68	96.22	6 , 3	8	✓		

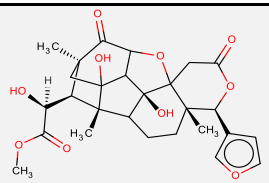
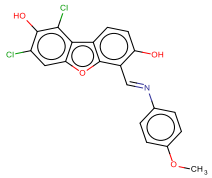
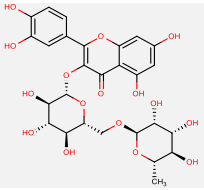
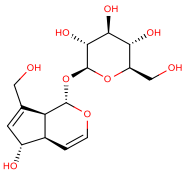
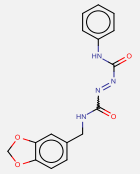
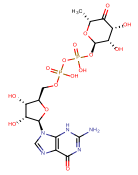
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.66, 83.01	ZINC15924502		429.73	7.06	32.67	2 , 0	3			
-9.66, 83.01	ZINC17353914		472.62	4.00	75.35	4 , 2	2	✓		
-9.66, 83.01	ZINC00032820		348.19	0.31	102.18	6 , 5	3	✓		
-9.66, 83.01	ZINC03812006		328.30	-0.69	130.27	5 , 1	6	✓		
-9.65, 84.42	ZINC09230252		400.47	-1.25	95.02	6 , 3	3	✓		

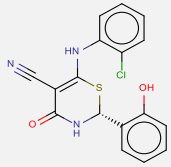
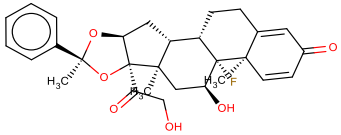
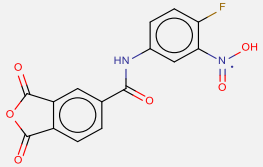
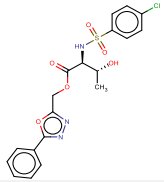
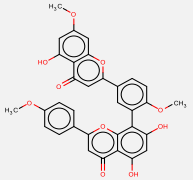
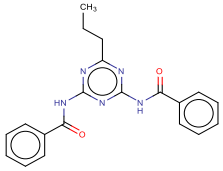
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.65, 84.42	ZINC29589828		396.91	4.54	35.94	4 , 1	6	✓	✓	
-9.65, 84.42	isokigelinol		308.33	1.83	74.6	4 , 2	0	✓		
-9.65, 84.42	cosmossin (acacetin-7-O-glucoside)		432.38	0.44	166.14	10 , 6	4			✓
-9.64, 85.86	ZINC01578220		447.50	4.56	81.71	5 , 2	6	✓		
-9.64, 85.86	ZINC04918956		516.45	2.16	211.28	10 , 7	9			

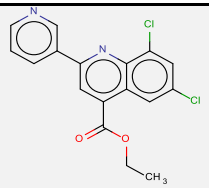
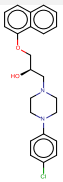
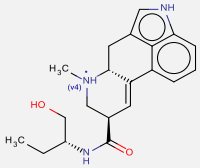
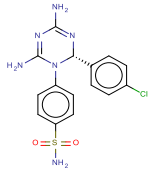
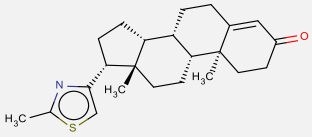
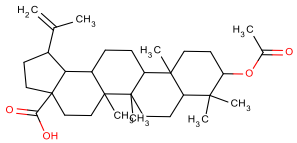
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.64, 85.86	*ent*-dioncophylleine A		377.48	3.80	50.72	4 , 2	2	✓		
-9.64, 85.86	1-(3-(2-ethoxyphenyl)ureidoacetyl)-4-(2-methyl-5-nitrophenyl)semicarbazide (CHEMBL566045)		442.43	2.45	171.57	6 , 5	8	✓		
-9.63, 87.32	ZINC13154300		369.54	3.81	40.54	3 , 1	1	✓		
-9.63, 87.32	ZINC02134706		562.63	0.37	190.36	9 , 4	10			
-9.63, 87.32	ancistrocladidine		405.49	4.60	60.28	5 , 1	3	✓		

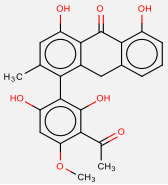
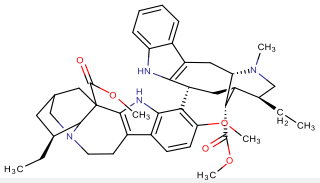
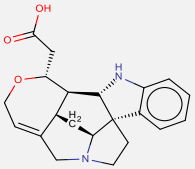
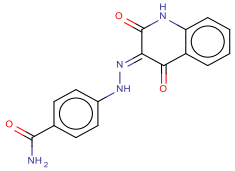
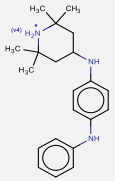
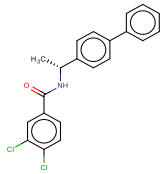
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.62, 88.81	ZINC13154304		363.54	5.21	36.02	1 , 2	0			
-9.62, 88.81	ZINC16951318		402.90	4.74	105.31	7 , 3	7	✓		
-9.62, 88.81	ZINC17465958		429.92	6.62	47.03	3 , 0	4			✓
-9.61, 90.32	ZINC13597432		342.39	2.42	78.53	4 , 1	4	✓		
-9.61, 90.32	acetylharpagide		406.38	-3.79	175.37	10 , 6	5			
-9.61, 90.32	praecansone B		366.41	3.72	64.99	5 , 1	5	✓		

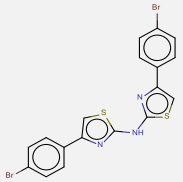
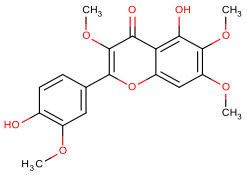
Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.6, 91.86	ZINC00001087		348.35	1.22	79.73	4 , 1	1	✓		✓
-9.6, 91.86	ZINC01736228		392.45	6.01	42.52	4 , 2	3			
-9.59, 93.42	ZINC00154888		302.32	3.67	54.37	3 , 1	4	✓		
-9.59, 93.42	ZINC08648354		394.81	4.29	90.13	5 , 1	5	✓		✓
-9.59, 93.42	ZINC13145028		386.48	0.95	77.95	5 , 3	3	✓		
-9.59, 93.42	dioncophylline B		363.45	3.51	61.72	4 , 3	2	✓		✓

Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.59, 93.42	1-O-deacetylkhayanolide E		516.54	0.47	152.73	7 , 3	4			
-9.57, 96.63	ZINC17949075		402.23	5.47	75.19	4 , 2	3			✓
-9.57, 96.63	rutin		610.52	-0.87	265.52	16 , 10	6			✓
-9.56, 98.27	aucubin		346.33	-3.18	149.07	9 , 6	4			
-9.56, 98.27	CA60		326.31	1.97	101.38	6 , 2	3	✓	✓	✓
-9.56, 98.27	GDP-4-keto-6-deoxy-d-mannose		587.33	-4.12	304.04	16 , 8	8			

Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.55, 99.94	ZINC01590864		357.81	3.58	85.15	4 , 3	3	✓		
-9.55, 99.94	ZINC11677178		496.57	3.65	93.06	6 , 2	3	✓		
-9.55, 99.94	ZINC17465972		331.23	1.18	128.98	6 , 2	3	✓		
-9.54, 101.65	ZINC02636757		451.88	1.77	131.62	6 , 2	8	✓	✓	
-9.54, 101.65	sciadopitysin		580.54	5.53	140.98	10 , 3	5			✓
-9.53, 103.38	ZINC01584497		361.40	4.75	96.87	5 , 2	6	✓	✓	

Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.53, 103.38	ZINC01622198		347.20	4.51	52.08	3 , 0	4	✓		
-9.53, 103.38	ZINC29589831		396.91	4.54	35.94	4 , 1	6	✓		
-9.52, 105.14	ZINC03831068		340.44	-2.52	69.56	3 , 4	4	✓	✓	
-9.52, 105.14	ZINC04824902		378.84	1.77	140.16	7 , 3	3	✓	✓	
-9.52, 105.14	ZINC17375647		369.56	4.93	29.96	2 , 0	1	✓		✓
-9.52, 105.14	acetylbetulinic acid		498.74	7.08	63.6	3 , 1	4			

Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.52, 105.14	knipholone anthrone		420.41	5.85	124.29	7 , 4	2			✓
-9.52, 105.14	tabernaelegantine A		706.91	6.53	99.89	5 , 2	8			✓
-9.51, 106.92	ZINC01711028		352.43	-1.66	61.8	5 , 2	2	✓		
-9.51, 106.92	ZINC18154478		308.29	2.86	113.65	5 , 3	3	✓		✓
-9.5, 108.74	ZINC00086102		324.48	-0.64	40.67	3 , 3	4	✓		
-9.5, 108.74	ZINC06343280		370.27	6.04	29.1	1 , 1	4			✓

Binding energy (kcal.mol-1), Ki (nM)	Name	Structure	MW	LogP	TPSA	H-bond Acceptors , Donors	Number of Rotatable Bonds	Lipinski Rule of Five Compliant?	Hit in VS1 ?	Hit in VS3 ?
-9.5, 108.74	ZINC13154298		493.24	7.70	37.81	3 , 1	4			✓
-9.5, 108.74	chrysosplenetin		374.34	2.40	103.68	8 , 2	5	✓		

7.2 The development of products for neglected tropical diseases

7.2.1 Foreword

The purpose of this chapter is to detail the various ways in which drugs and vaccines for neglected tropical diseases (NTD) are developed and advanced towards the market. This chapter will explore factors affecting product (drug/vaccine) development for NTDs and will indicate how they apply to the development of a drug to treat human African trypanosomiasis (HAT) or Chagas Disease. There are a few excellent resources that inform on the funding available for NTD product development. One such resource is the Global Funding of Innovation for Neglected Diseases (G-FINDER) report(1) and database(2), the results of a survey on funding for NTD research & development (R&D), compiled by Policy Cures. Another useful resource is BIO Ventures for Global Health's (BVGH) Global Health Primer(3), a database that holds details on NTD products currently in development and the organisations involved in those development projects. In addition to these, the research underlying this chapter was guided by a number of reviews of NTD product development activity(4–7). However, there appears to be a gap in the research, mainly pertaining to the different strategies used by researchers and developers to advance products towards the market and the availability of funding to advance development at specific stages in the pharmaceutical research and development (R&D) lifecycle. Furthermore, the literature suggests that the formation of partnerships between different organisations is almost ubiquitous amongst NTD product development campaigns, although there is a scarcity of reviews providing specific details on such partnerships. Therefore, this chapter will cover the following areas: the different commercialisation strategies for new products to control NTDs, the level of evidence required when entering into partnerships, and the nature of partnerships formed during pharmaceutical development for NTD products. The implications for NTD product developers are presented from the point of view of researchers in academic and public or not-for-profit research institutions (PSNRI), and recommendations on the most appropriate strategies to be utilised in the pursuit of the research discussed in this thesis will be given.

7.2.2 Introduction

7.2.2.1 The NTD health burden

NTDs collectively represent an enormous health burden globally. These diseases are infectious diseases, generally concentrated in the Global South. The WHO has defined a list of 17 NTDs(8), whilst a number of other diseases have been identified as NTDs by other organisations(9,10) (see Figure 7.2.1). It has been estimated that NTDs represent 10-12% of the global disease burden(4,5,11), but receive a relatively small investment in R&D. NTDs are now often referred to as diseases of "neglected populations", as they often exist in communities that are socially excluded, sometimes due to their geographic location, but also as a result of their impoverished condition, to which infectious diseases contribute(9). As a result of this situation, NTDs lie as a barrier to economic development and represent a significant expenditure to governments in disease-endemic areas (9). For example, an estimated US$267million is spent annually on the treatment of Chagas Disease

patients in Colombia alone(12). Thus with such economic and social costs resulting from NTDs globally, it is within the interests of private, public and philanthropic organisations, both nationally and internationally, to bring them under control.

"The Big Three"
- HIV/AIDS
- Malaria
- Tuberculosis

Protozoan Infections
- Amebiasis
- Balantidiasis
- Chagas Disease*
- Giardiasis
- Human African Trypanosomiasis*
- Leishmaniasis*

Helminth Infections
- Taeniasis-Cysticercosis*
- Dracunculiasis*
- Echinococcosis
- Food-borne Trematodiases*
- Loiasis
- Lymphatic Filariasis*
- Onchocerciasis*
- Schistosomiasis*
- Soil-transmitted Helminthiases (Ascariasis, Hookworm Diseases, Trichuriasis, Strongyloidiasis)*
- Toxocariasis and other Larva Migrans

Other
- Snakebite

Viral Infections
- Dengue*
- Japanese encephalitis
- Jungle yellow fever
- Other arboviral infections
- Rabies*
- Rift Valley fever
- Viral hemorrhagic fevers

Bacterial Infections
- Bartonella
- Bovine Tuberculosis in Humans
- Buruli Ulcer*
- Cholera
- Enteric pathogens (Shigella, Salmonella, E. coli)
- Leprosy*
- Leptospirosis
- Relapsing Fever
- Trachoma*
- Treponematoses (Bejel, Pinta, Syphilis, Yaws*)

Fungal Infections
- Mycetoma
- Paracoccidiomycosis

Ectoparasitic Infections
- Scabies
- Myiasis

Figure 7.2.1 - Diseases classified as NTDs. This list was constructed based on definitions by Hotez et al.(9) and the scope of the journal PLOS Neglected Tropical Diseases(10). "The Big Three" are excluded from some definitions in the literature, as they receive much more attention than other NTDs. * indicates a disease included in the WHO list of 17 core neglected tropical diseases.

7.2.2.2 NTD policy

Historical

In the disease endemic countries, many NTDs have been present for several thousands of years, periodically presenting as epidemics with several fatalities. The human and economic costs of NTDs were recognised by the colonial powers of former colonies, which led to the implementation of mass screening, treatment and vector control problems(4,13). In fact, many of the most burdensome NTDs once posed barriers to economic progress in Western countries, and their subsequent control and

elimination increased the prosperity of affected areas. However after several of the former colonies gained independence and disease prevalence had decreased, control programs were relaxed and some of these diseases began to re-emerge. Since that time, various famines, regional conflicts and political crises have exacerbated the situation.

Awareness and policy

There are a number of reasons for the persistence of today's NTDs. The WHO have described how certain diseases lack the appropriate "tools" (e.g. drugs, vaccines, diagnostics) to control them, whereas with other NTDs, we are limited by a lack of understanding by the disease pathology(14). However, the reason most commonly cited as a cause of the "market failure" to encourage NTD research and product development, is their association with poverty. Patient populations lack purchasing power, thus, there is little or no financial incentive for big pharmaceutical companies, the traditional developers of new drugs and vaccines, to create new treatments(4,5,7,15). Consequently, large Pharmaceutical corporations have shied away from NTD product development, with research into new treatments for NTDs once coming to a virtual standstill(4). For certain NTDs, this has prepared the way for a possible health crisis, as resistance to old drugs emerges and disease prevalence increases.

The need for strong policy decisions and national and international governance to support the development of new NTD products has been argued for by several proponents(7,15). Unfortunately, for a time, programs to control/eliminate NTDs were deficient, and NTDs were not formally recognised by health organisations – the WHO originally failed to specifically mention several NTDs in their Millennium Development Goals(16). In recent years, thanks to the restless awareness-raising activities of various researchers and NGOs, the fight against neglected diseases is now supported by clear policies in a number of institutions, globally. This change has been accompanied by unprecedented structural changes in the pharmaceutical industry which have helped to spur the development of a number of new drugs and vaccines for NTDs(7,17).

Recent advances

Some of the policies designed to spur NTD product R&D can be categorised as "push" and "pull" mechanisms. "Push" mechanisms are those that aim to encourage or enable R&D by cutting the costs of doing so. Product researchers and developers usually gain access to these through collaborations with governmental organisations and charities. "Pull" mechanisms encourage investment in R&D by creating financial incentives for this activity. Examples of these are Advanced Market/Purchase Commitments (AMCs) and the FDAs Priority Review Vouchers (PRV)(6,18). Although the effectiveness of these "pull" mechanisms is yet to be proven(18–22), they do theoretically solve the problem of a deficient market to incentivize NTD product R&D. The "push" mechanisms have already proven their worth in spurring several new R&D projects, whilst also re-engaging pharmaceutical companies(5,7). The majority of this has happened against a backdrop of a

paradigm shift in pharmaceutical development for NTDs, the product development partnership (PDP)(7,23), which will be discussed in detail below.

Although still much smaller than for other therapeutic areas, the global investment in pharmaceutical R&D for NTDs has increased vastly. In 1999, it was estimated that less than US$70 million were invested into drug research for malaria, tuberculosis, HAT and leishmaniasis, combined. In contrast, the 2012 G-FINDER report stated that US$ 3,045 million were invested globally in NTD research, with around 70% of that directed towards product development (as reported by respondents to the G-FINDER survey)(1). Although funding, particularly from philanthropic and public donors, has declined since the global financial crisis, funding levels are still higher than in 2007(1), and much greater than 15 years ago(4).

Many national governments are now some of the biggest funders of NTD R&D. Most prominent are the US, through the US Agency for International Development (USAID), and the UK, through its Department for International Development (DFID). Other countries, such as The Netherlands and Australia, have emerged as major NTD R&D funders(1). Meanwhile, the WHO's Special Program for Research and Training in Tropical Diseases (WHO TDR) continues its work with a wide range of private and public R&D partners around the world(1,24,25).

7.2.2.3 The Product Development Partnership (PDP) model

PDPs bring together several different partners with the goal of developing one or a related group of products. PDPs are typically partnerships between public and private institutions, where the relative strengths of each individual institution can be exploited. This is particularly useful in developing products for neglected tropical diseases, for a number of reasons. Philanthropic organisations are able to allocate resources to the researchers and developers, covering or reducing the costs of R&D. Academic and other public-sector or non-profit research institutions (PSNRIs) are very-well placed to carry out discovery and early-stage development work(26–28), whilst biotechnology and pharmaceutical firms are well technologically resourced. Large pharmaceutical firms have a wealth of experience in conducting clinical trials, but this can also be supplemented by the knowledge and experience of institutions and NGOs operating in NTD-endemic countries. This ability to leverage the strengths of each partner has likely contributed to the success of the PDP model. It has resulted in clinical development times that are comparable with more profitable pharmaceuticals. Also, average development costs are significantly cheaper than for the development of more profitable drugs, outside of the PDP model(7,29).

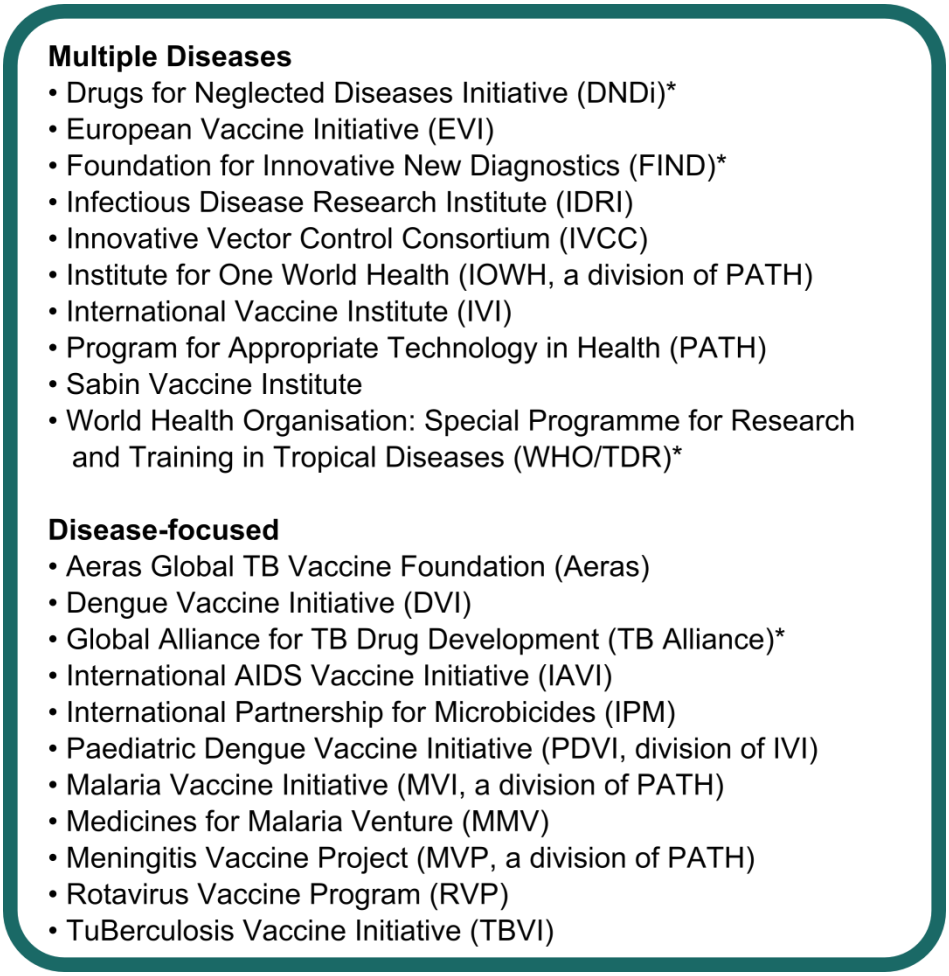


Figure 7.2.2 - examples of PDP "Brokers" involved in product development for drugs, vaccines, diagnostics and insecticides for NTDs. * indicates that they participate in the development of products to control trypanosomiasis(2,3,29).

Several of the PDPs are co-ordinated by “Brokers”, such as the Drugs for Neglected Diseases Initiative (DNDi) and BIO Ventures for Global Health (BVGH). These partnership brokers are important development partners, due to their extended networks, experience and central role in allocating funding. Disease-specific alliances (e.g. The Medicines for Malaria Venture, MMV) have also emerged as important brokers of PDPs and as developers in their own right.

7.2.2.4 Imbalanced R&D priorities

Despite these positive changes, a number of concerns in NTD product development still remain. One is the gaps in research, which stem from the lack of sustainable funding for some diseases. This problem will be minimised if global funding increases, and if partnerships continue to be formed appropriately and managed effectively(7). What may potentially be a bigger problem, from a global health perspective and from a researcher’s perspective, is the bias of product development for re-formulations and fixed dose combinations, rather than for new chemical entities(5)(see Table 7.2.2, below). Trouiller and colleagues describe how the lack of appropriate “tools” to treat certain NTDs, requires that innovative new products are developed – new chemical entities (NCE)s and novel

vaccines. However, for drugs in general, and to a greater extent for NTD drugs, very few NCEs are developed when compared to other registered products(5,11).

7.2.3 Trypanosomiasis

The two human diseases caused by trypanosome infections, human African trypanosomiasis (HAT) and Chagas Disease, are classic examples of NTDs. They are debilitating and fatal diseases that primarily affect the poor. HAT, which is prevalent in Sub-Saharan Africa, and Chagas Disease, which is concentrated in Latin America, both have a long history. Recent efforts have led to improved control of both disease, but problems persist.

There has been a significant decrease in the number of new cases of HAT(30). Effective control programs led to a reduction of 68% between 1995 and 2006. 36 countries remain HAT-endemic, although only seven of these have reported more than 100 cases per year in the period 1997-2006(13). However, given the situation in those countries of higher prevalence, along with the ever-present possibility of HAT re-emergence among an at-risk population of 70 million(30), it is important that there are appropriate tools at the disposal of health authorities.

Although control programs (involving screening, vector control and treatment programs) have proven very effective in reducing the number of HAT cases, the effect has been greater for one form of HAT (the *gambiense* form), because an animal reservoir plays a significant role in the spread and persistence of the other (*rhodesiense)* form(13,30).

Due to the chronic nature of Chagas Disease, diagnosis is challenging and infection often goes unnoticed for several years. Between 7 and 8 million are estimated to be infected with Chagas disease worldwide, and it is spreading beyond its original geographical location to other continents. The treatments are effective if used shortly after infection. However, treatment regimens are long, and adverse reactions occur in up to 40% of patients. The majority of patients will present with chronic infection, often with other complications, when the effectiveness of the usual treatments is limited.

Control of the Chagas Disease-spreading vector, the triatomine bug is probably the most effective means of controlling Chagas Disease, although there are challenges of the re-emergence of the disease in certain areas and the need for more effective treatments to treat chronic infection(12).

7.2.4 Trypanosomiasis R&D

Shortfalls

The biggest shortfall in the control of HAT lies in the tools that are currently available. A HAT vaccine is very challenging, due to the biology of the pathogen(31,32), so efforts are focused on vector control and treatments. All treatment regimens involve some form of injection, which requires more resources and well-trained staff. The length of some of these regimens puts a further strain on resources, which are often not well-developed in rural areas of developing countries. The current

drugs all suffer from a high incidence of side-effects, which can be burdensome or even fatal in some cases. Therefore, there is an urgent requirement for new drugs, with less side-effects and more suitable formulations(13,33).

Chagas Disease treatments suffer from similar drawbacks to those of HAT treatments. As stated, above, a new drug for Chagas disease should ideally be more effective than current treatments in chronic stages of the disease and should have a lower incidence of side effects.

7.2.5 Drivers of NTD R&D

The main drivers of NTD product R&D come from philanthropic organisations, the pharmaceutical industry, governments, PDP brokers and other research institutions (academic and PSNRI). Prior experience in NTD product R&D strongly indicates that partnering with one or several of these organisations would be imperative. This chapter focuses on NTD development from the point-of-view of researchers from an academic institution/PSNRI. Hence, the following discussions will analyse different commercialisation strategies and potential partners from that viewpoint.

Groups from Universities or PSNRIs play a very important role in the discovery of new innovative treatments(26,34), but they also have several limitations as projects progress towards the market(27,35,36). Therefore, partnerships are essential to them.

7.2.6 Commercialisation strategies

7.2.6.1 General strategies for pharmaceutical products

The first step in commercialising scientific research typically entails the generation and protection of intellectual property (IP) pertaining to an invention. With pharmaceuticals, this will usually involve filing a patent. The holder of the patent can then seek to capitalise on this IP through licensing of the invention to a third party, who may have more resources to develop and distribute a marketable product. Another strategy to reap rewards from IP is to form a spin-out company. Funds to develop the invention are raised in the form of equity in the company. The company may enter into a partnership with third parties to share risk and to co-develop the product. The university spin-out company is increasingly becoming a method favoured by Universities to exploit new IP and this plays an important role in new product discovery and development(37). This situation holds true in new drug and vaccine development by Universities(27) and companies alike. Both academic institutions and new companies have been identified as important sources of new drugs(26,34).

Given the high risks and capital costs involved in pharmaceutical R&D, the formation of co-development partnerships is almost ubiquitous. Large pharmaceutical corporations have looked to academic research institutions and small/medium biotechnology or pharmaceutical enterprises (SMEs) as sources of new drugs and vaccines(28). Meanwhile, their smaller, less resourced partners are able to benefit from the extensive capabilities and product development experience of the larger

firms. Certain R&D activities are often outsourced to contract research organisations (CROs), often as a cost saving measure, but sometimes as a faster and more efficient way to advance products through clinical development. When new product development is initiated by small companies, and sometimes when large companies are the sole developer, it may be necessary to bring in a number of other partners (venture capitalists and other shareholders) to finance development.

The following section discusses how different development partners may be involved in the development of drugs for a neglected tropical disease, such as trypanosomiasis.

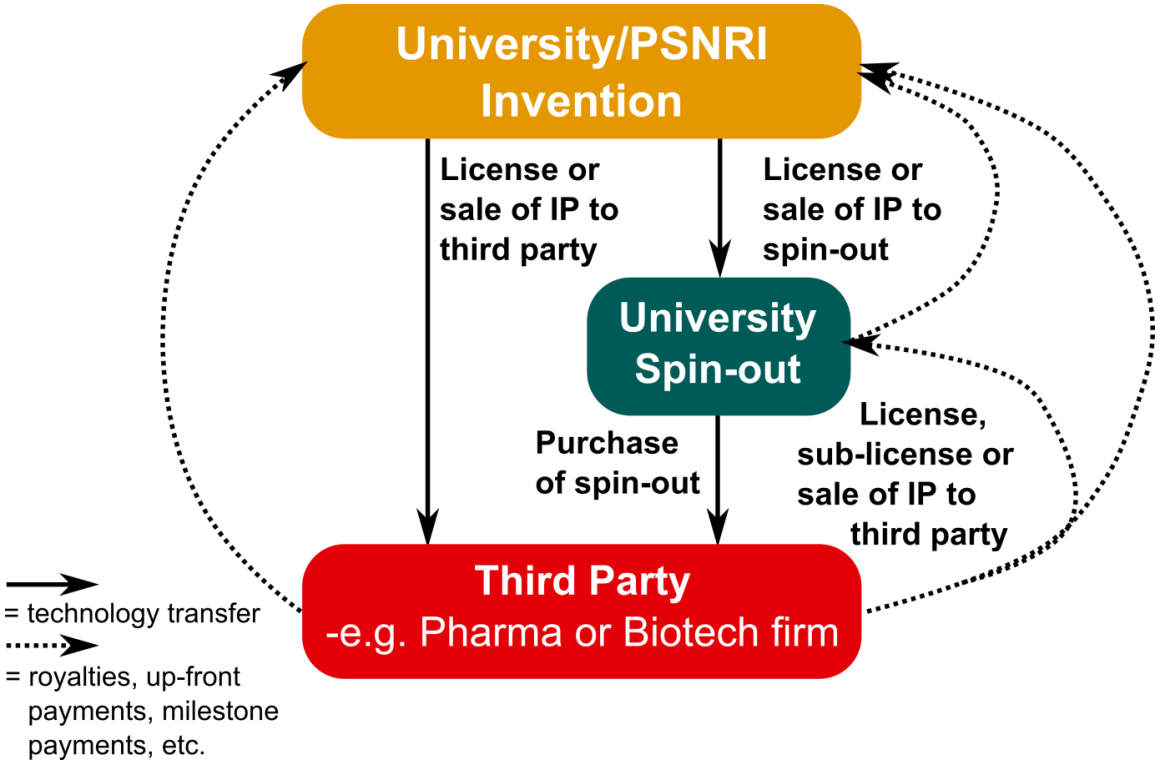


Figure 7.2.3 - scheme illustrating potential commercialisation strategies for university- or PSNRI-originating research. IP is either licensed/sold directly to third parties, or to a spin-out company formed by the inventors' institution. Solid arrows indicate technology transfer, whilst dotted lines indicate returns to the University/PSNRI.

7.2.7 Product development partners

7.2.7.1 The general picture

NTD product development inherently faces the same risks as the development of developed-world diseases. However, as previously mentioned, there is a huge disparity between the collective or individual disease burdens of NTDs and the R&D activity aimed at controlling them (4). Despite this, the funding for global NTD product R&D has increased dramatically over the previous decade, showing gradual increases up until 2009, and steady levels since then. 71.4% of the US $3.045 billion NTD R&D funding in 2011 was in the form of external grants, with a reported $1.5 billion going to academic and private NTD product developers(1).

The largest funders of NTD research are government institutions, presenting as aid agencies or research councils/science and technology (S&T) agencies. They provide around 65-70% of global funding for NTD research. The other two major sources of funding for NTD research are industry (pharmaceutical and biotechnology) and philanthropy, contributing 17.2% and 18.7%, respectively, of all global funding in 2011(1).

PDPs take a central role in NTD drug development. They received 23% of global grant funding for NTD research in 2011(1). Furthermore, the BVGH 2012 report on the product developer landscape showed that 40% of all drugs and vaccines in current development were being developed using the PDP model(38). Hence, the PDP brokers are another key partner in NTD product development.

7.2.7.2 Funding for trypanosomiasis

4.3% (US $131.7 million) of global funding reported in the G-FINDER report was dedicated to research in diseases caused by kinetoplastids (leishmaniasis, HAT and Chagas disease). Of this amount, 26% was granted to HAT and 18.7% to Chagas disease. A total of roughly US$18 million was spent on R&D for new drugs in 2011(1). The funding landscape for these diseases appears close to the general picture: 73.3% of funding originates in the public sector and 17.3% is donated by philanthropic organisations. The top three funders of research in kinetoplastid diseases are the US National Institutes of Health (NIH), The Bill & Melinda Gates Foundation (Gates Foundation) and The Wellcome Trust. Industry plays a smaller role here, than in other disease areas, contributing only 9.5% of funding, but it is still the third largest source of funding.

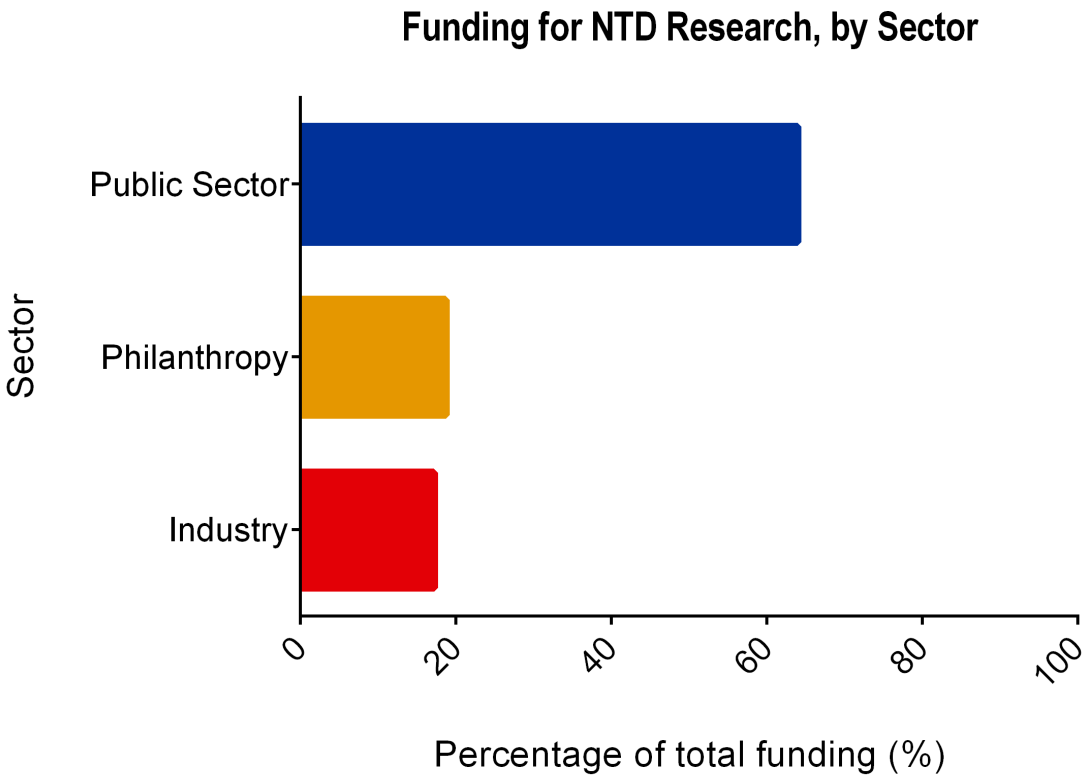


Figure 7.2.4 - global funding for NTD research in 2011, as reported in the G-FINDER report(1). Funding by sector is presented as percentage of total funding.

7.2.7.3 The academic/PSNRI institution as a source of new products for NTDs

In all disease areas, academic institutions and PSNRIs have proven to be important sources for new drugs and vaccines(26). Analysis of some recently-registered NTD products shows academics and PSNRIs at the root of their discoveries(4). A look at the BVGH Global Health Primer also reveals the involvement of such institutions in the development of several NTD products in the pipeline(3). This role in product R&D is particularly important in the NTD setting; the development of innovative treatments is an inherent requirement to treat or prevent several NTDs(4). Evidence to support an innovative role can be found in the literature. In Stevens et al.'s analysis of new drug applications to the US Food and Drug Administration (FDA), they show that academics/PSNRIs were twice as likely as for-profit institutions to meet an unmet medical need when a new drug originated with them. Academic/PSNRI inventors also contributed to 13.6% of new molecular entities (NMEs) included in the analysis, and 9 out of the 10 new drugs that were for new indications(26). Similarly, in Kneller's 2010 analysis of the origins of new drugs, Universities were shown to contribute to the R&D of 30% of drugs given the FDA's priority review (see discussion of PRV, below) and to 31% of drugs considered to be scientifically novel, despite contributing to an estimated 24% of new drugs overall(34).

The prominent role of public funders, coupled with the fact that around three-quarters of all public funding goes directly to researchers(254), means that academic and PSNRIs are collectively the largest recipients of NTD R&D funding. There has also been an increase in intramural funding by PSNRIs and Universities(1). This makes researchers in these positions the most common

participants in NTD product development(38). However, groups from Universities and other PSNRIs are often insufficiently resourced to progress new products towards the clinic alone. The costs of preclinical drug development can cost hundreds of thousands of pounds at a minimum, which can be very inhibiting, particularly to a small academic group(36). On the other hand, with the rise of Academic Drug Discovery Centres, particularly in the US, more academic and PSNRI researchers will find the necessary resources available to them(27). Even small groups can collaborate with other groups within their institution or externally in order to carry out different studies using the grants of individual researchers. When looking at the current NTD product pipeline, academic/PSNRIs can be seen to be involved at every stage of the product development lifecycle – from discovery through to clinical phases(3,38).

An academic institution/PSNRI has the option to sell the IP of a particular invention to another product developer to take forward. This is clearly the most convenient option and, in many cases, it is the most appropriate option, relieving researchers of duties that are outside of their area of expertise. There are situations where the development of a product will benefit from the continued input of the original inventors, in which case it would be a good option to license the product to third parties and co-develop the product(39). In addition to these strategies, academics/PSNRIs may spin-out a company to commercialise a product. This has the added value of being able to attract more funds in the form of equity. Being able to provide more funds directly, means that the inventors take on more of the risk, but are in a more favourable position to reap the commercial rewards of any products that meet the market. Thus, the decision to follow either of the last two strategies is largely dependent on the commercial value of the product being developed. Universities often pursue the spin-out model as a more lucrative revenue source(28,39). However, in the NTD space, most products would have little or no commercial value. This is particularly the case for HAT and it is probably the case for Chagas disease.

The contribution of academic/PSNRIs in overall drug discovery and development, beyond the discovery phase is confirmed by numerous sources(26–28,34). However, in the vast majority of cases, there is some form of technology transfer from Universities/PSRIs to Pharmaceutical or Biotechnology companies, who complete late-stage development, and carry out registration, marketing and distribution. In Kneller's study of the origins of drugs, 8% of drugs originated from Universities that transferred their technology to Pharmaceutical companies, whilst 16% originated from Universities that transferred their technology to Biotechnology companies.

Table 7.2.1. Organisations engaged in the development of drugs to treat trypanosomiasis. Organisations were identified as funders of drugs for trypanosomiasis who reported to the G-FINDER survey(2), or as participants in current projects to develop trypanosomiasis drugs on the BVGH global health primer(3). Some other organisations involved in NTD product R&D were identified through media reports or were signatories of the London Declaration to combat NTDs(40). This list is not exhaustive. In particular, the PSNRI included here probably represent only a small sample of all those that are involved in this area. Several of these organisations are potential partners, whereas others, such as foreign government departments, will be funders of these partners. The capabilities here are the main capabilities of each organisation in this area and were judged based on reports of past activity in trypanosomiasis or NTD drug development and on information posted on the organisations' websites. *SMEs are pharmaceutical or biotechnology companies with up to 250 employees. Some of the PSNRIs, such as the Institute for One World Health and Seattle Biomed, have similar capabilities to the SMEs listed here.

Name	Type	Capabilities	Experience with HAT/Chagas Disease?	Comments
Abbvie Inc.	Pharmaceutical Company	R&D: entire lifecycle		A division of Abbott
Anacor Pharmaceuticals	SME*	R&D: entire lifecycle	✓	Has required partners to develop systemic products past preclinical stage
aRigen Pharmaceuticals	SME	R&D: discovery - phase I	✓	
AstraZeneca	Pharmaceutical Company	R&D: entire lifecycle		
Barcelona Centre for International Health Research	PSNRI	R&D: discovery - preclinical	✓	
Bayer Healthcare	Pharmaceutical Company	R&D: entire lifecycle		
Belgian National Fund for Scientific Research (FWO)	S&T Agency	Funding	✓	
Bill & Melinda Gates Foundation	Philanthropy	Funding; Partnership	✓	
Biotechnology and Biological Sciences Research Council (BBSRC)	S&T Agency	Funding	✓	

Name	Type	Capabilities	Experience with HAT/Chagas Disease?	Comments
Brazilian Development Bank (BNDES)	Government	Funding	✓	
Brazilian Foundation Araucaria, Fundaçao Araucaria	Government	Funding	✓	
Brazilian Ministry of Health: Department of Science and Technology (DECIT)	S&T Agency	Funding	✓	
Brazilian Research Support Foundation of the State of Bahia, Fundaçao de Amparo a Pesquisa do Estado da Bahia (FAPESB)	S&T Agency	Funding	✓	
Canton of Geneva	Government	Funding	✓	
Carlos III Health Institute	PSNRI	R&D: discovery - clinical trials	✓	
CDC, Centre for Global Health, Division of Parasitic Diseases and Malaria	PSNRI	R&D: discovery - preclinical		
Centre for Discovery and Innovation in Parasitic Diseases (Sandler Centre for Drug Discovery)	University/PSNRI	R&D: discovery - preclinical	✓	
Developing World Health	PDP broker	Partnerships; Implementation		
Drugs for Neglected Diseases Initiative (DNDi)	PDP broker	Partnerships	✓	
Dutch Ministry of Foreign Affairs - Directorate General of Development Cooperation (DGIS)	International Aid	Funding	✓	
Eisai Company	Pharmaceutical Company	R&D: entire lifecycle	✓	
Eli Lilly & Company	Pharmaceutical Company	R&D: entire lifecycle		
Epichem	CRO	R&D: discovery, pharmaceutical formulation	✓	

Name	Type	Capabilities	Experience with HAT/Chagas Disease?	Comments
European Commission: Directorate-General for Research and Innovation	Government	Funding	✓	
French Ministry of Foreign and European Affairs, Ministere des Affaires Etrangeres et Europeennes (MAEE)	International Aid	Funding	✓	
Genomics Institute of the Novartis Research Foundation	Pharmaceutical Company	R&D: discovery - preclinical	✓	A subdivision of Novartis
German Agency for Technical Cooperation (GTZ) Deutsche Gesellschaft für Technische Zusammenarbeit GmbH	Government	Funding	✓	
Gilead Sciences	Biotechnology Company	R&D: entire lifecycle		
GlaxoSmithKline	Pharmaceutical Company	R&D: entire lifecycle		
Global Alliance for TB Drug Development	PDP Broker	Partnership; Funding	✓	

Name	Type	Capabilities	Experience with HAT/Chagas Disease?	Comments
Global Health Innovation Technology Fund	PDP Broker	Partnership; Funding		Non-profit founded by Japanese firms (Astellas, Daiichi-Sankyo, Eisai, Shionogi, Takeda), Japanese Government, UNDP and The Gates Foundation
Global Network	Philanthropist	Implementation; Advocacy		
Immtech Pharmaceuticals	SME	R&D: discovery - preclinical	✓	
Indian Department of Biotechnology, Ministry of Science and Technology (DBT)	S&T Agency	Funding	✓	
Infectious Disease Research Institute	PSNRI; PDP Broker	R&D: entire lifecycle; Partnerships		
Institut Pasteur	PSNRI	R&D: discovery - clinical trials	✓	
Institut Pasteur, Korea	PSNRI	R&D: discovery - preclinical	✓	
Institute for One World Health	PSNRI; PDP Broker	Partnership; R&D: discovery - clinical trials		A subdivision of PATH
Inter-American Development Bank	Bank	Funding		

Name	Type	Capabilities	Experience with HAT/Chagas Disease?	Comments
Janssen Pharmaceutical Companies	Pharmaceutical Company	R&D: entire lifecycle		A subdivision of Johnson & Johnson
Johnson & Johnson	Pharmaceutical Company	R&D: entire lifecycle		
Malaysian Ministry of Science and Technology (MOSTI), including the National Biotechnology Division (BIOTEK)	S&T Agency	Funding	✓	
MANLAB	CRO	R&D: preclinical		
Médecins Sans Frontieres/Doctors Without Borders (MSF/DWB)	NGO	Funding; Implementation	✓	
Merck & co.	Pharmaceutical Company	R&D: entire lifecycle	✓	
Mundo Sano	PDP Broker; PSNRI	Partnership		
National Program to Fight HAT, DRC	Government	Implementation	✓	
National Scientific and Technical Research Council - Argentine (CONICET)	S&T Agency	Funding	✓	
Novartis AG	Pharmaceutical Company	R&D: entire lifecycle		
Novartis Institutes for Biomedical Research (NIBR)	Pharmaceutical Company	R&D: discovery - preclinical		A subdivision of Novartis
Otsuka Pharmaceutical Company	Pharmaceutical Company	R&D: entire lifecycle		
Pfizer	Pharmaceutical Company	R&D: entire lifecycle		
Pharmaceutical laboratory of Pernambuco state (LAFEPE)	PSNRI	R&D: pharmaceutical formulation	✓	
Program for Appropriate Technology in Health (PATH)	PDP Broker; PSNRI	Partnership; R&D: discovery - clinical trials		
Roche	Pharmaceutical Company	R&D: entire lifecycle		

Name	Type	Capabilities	Experience with HAT/Chagas Disease?	Comments
RTI International	PSNRI	R&D: discovery - preclinical		
Sandoz Family Foundation	Venture Capital	Funding	✓	
Sanofi	Pharmaceutical Company	R&D: entire lifecycle	✓	
Sasakawa Peace Foundation	NGO	Funding	✓	
Save the Children	NGO; Philanthropist	Funding		
Scottish Higher Education Funding Council	Government	Funding	✓	
Scripps Research Institute	PSNRI	R&D: discovery - preclinical	✓	
Scynexis, Inc.	SME/CRO	R&D: discovery - preclinical	✓	
Seattle BioMed	PSNRI	R&D: discovery - clinical trials		
Spanish Ministry of Foreign Affairs and Cooperation for Development (MAEC) and/or Agency of International Cooperation for Development (AECID)	International Aid	Funding	✓	
Swiss Tropical and Public Health Institute (Swiss TPH)	PSNRI	R&D: discovery - clinical trials	✓	
Takeda Pharmaceutical Company	Pharmaceutical Company	R&D: entire lifecycle	✓	
The END Fund	Philanthropist	Funding		
The Wellcome Trust	Philanthropist	Funding	✓	
Therapeutics for Rare & Neglected Diseases (TRND)	PSNRI	R&D: preclinical - phase I		A subdivision of NIH
Tres Cantos Open Lab Foundation	PSNRI	R&D: discovery - preclinical		
UBS Optimus Foundation	Philanthropist	Funding	✓	Focus on Child well-being

Name	Type	Capabilities	Experience with HAT/Chagas Disease?	Comments
UK Department for International Development (DFID)	International Aid	Funding	✓	
UK Medical Research Council (MRC)	S&T Agency; PSNRI	R&D: discovery - clinical trials	✓	
US National Institutes of Health (NIH)	S&T Agency; PSNRI	R&D: discovery - clinical trials	✓	
WIPO Re:Search (BIO Ventures for Global Health)	PDP Broker	Partnership		Managed by BIO Ventures for Global Health

7.2.7.4 Partners for NTD product development

PDP Brokers

As previously described, the successes of several NTD drug and vaccine development programs have been attributed to the PDP model(29). The PDP brokers play a central role in this and manage much of the global funding that is directed to NTD research and development (see above). The characteristics of these brokers differ slightly. There are those that focus on a particular project, whilst there are others that are focused on a particular disease or group of diseases (e.g. Medicines for Malaria Venture [MMV], Global TB Alliance) and manage several products. Brokers such as DNDi and BVGH work across therapeutic areas in the NTD space. As the Brokers have extensive networks and are present at every stage of pharmaceutical development, it would be advantageous to engage a PDP early on in the process. The multi-project and multi-disease Brokers have the added advantage that they have more experience, as they tend to have existed for longer, and thus they are better connected. One disadvantage of PDPs is that they do not generate their own income and rely on funding from external sources. Theoretically, this would make them vulnerable to the same inconsistency in funding as academics/PSRIs. However, this does not seem to be the case, judging by the literature and media reports.

According to the BVGH Global Health Primer, DNDi and the Global Alliance for TB Drug Development, along with BVGH, are PDP brokers that are currently involved in trypanosomiasis drug development.

Government

The biggest proportion of global R&D funding for NTD research comes from government sources. A lot of this funding is given in pursuit of international development goals or for scientific reasons. According to the G-FINDER report, the US NIH (granting a huge US$1.16 billion over five years) is by far the most significant funder of NTD research. Researchers in the UK have access to funds from the European Commission (EC), UK DFID and The Medical Research Council, which are all in the top twelve funders of NTD research. Aid agencies and Research Councils/S&T agencies of the US, Dutch and Australian governments are also present in the list(1). There has been a shift over the years towards more basic research, with 31.2% of funding being dedicated to this area. Therefore, researchers are likely to find funds from these sources for discovery and early-stage development. However, a substantial proportion of funds are still available for later development. For example, in 2011-2012, the MRC released a Highlight Notice, prioritising translational research in NTDs(41). This remains a priority for the MRC's Infection and Immunity Board(42). Funds from research councils/S&T agencies are usually disbursed through competitive grant systems, where researchers must submit project proposals in response to calls for applications.

A large part of the support from government aid agencies focuses on implementation of mass treatment administration programs. This certainly seems to be the main role of USAID(43). The UK DFID is also heavily involved in this area, but also places a lot of emphasis on research to combat NTDs(44). Their contributions are delivered through the WHO TDR initiative and DNDi programs(45)

Philanthropy

Philanthropic donations for NTD research do not seem to be concentrated in any particular disease area. This funding usually plays a contributing role to specific projects and is focused on product development. Around half of all PDP funding comes from philanthropic organisations (principally The Gates Foundation)(1). Thus, it is important to highlight that, although Philanthropists are not as prominent as governments in NTD research funding overall, they are often major sources of funding for the development of individual products. The two largest philanthropic funders are The Gates Foundation, providing an average of US$120.3 million annually in 2007-2011, and The Wellcome Trust, which gave an average of US$10.8 million annually to NTD research over the same period(1). The main areas of focus of The Gates Foundation are Global Health, Global Development, Education (in the US) and Global Policy & Advocacy(46). Efforts to control NTDs receives strong support from the Gates Foundation's Global Health Division and HAT is one of the "High Opportunity Targets" that they focus on(47). Apart from funding PDPs and other global health organisations, The Gates Foundation also has a mechanism for giving grants to researchers. Concepts with specified aims are developed internally, then researchers/organisations are either invited privately or publically to apply for grants in a similar way to how they would apply for research council grants(48). The Wellcome Trust operates in a similar way to research councils. Infectious Diseases are included in the list of 'Research Challenges' in the Trust's 2012-2020 Strategic Plan(49). There are also two specific programs which are applicable to NTD product R&D. Firstly, the Global Health trials scheme provides funding for late-stage clinical trials of 'interventions that will improve health' in low- and middle-income countries(50). Secondly, the Wellcome Trust's Pathfinder Awards provide funding for Academic-Industry partnerships for R&D for NTDs and Orphan Diseases. The aim of this initiative is to advance projects from an early-stage to a stage where there is enough credibility to attract funds from other sources(51). Target validation is not supported by this scheme, so it can be envisaged that this scheme would be attractive to researchers working in the discovery and pre-clinical development stages. Furthermore, similar lines of research could be funded through other, less specific Wellcome Trust Grants.

There is little information on the role of Philanthropy by wealthy individuals in funding NTD research. It is possible that these individuals will give their donations to PDP brokers, and other charities/non-profit organisations, so their contribution is unclear. Given the engagement of such individuals in issues of social importance(52), they should be approached to support R&D for NTD product development.

Large Pharmaceutical Companies

Pharmaceutical firms are one of the primary partners of academic/PSNRI organisations for drug and vaccine development. They are also a key partner in NTD drug development. A search of the G-FINDER database, and a review of the signatories to the London Declaration on NTDs reveals a significant number of large pharmaceutical corporations that are involved in NTD product R&D(1,40). They are responsible for a large part of the industry participation in NTD product development and some companies are engaged in the development of several products at a time(38). In addition to the potentially large amount of funds at their disposal, large pharmaceutical firms also contribute to a large extent with in-kind support, ranging from access to facilities/compounds, offering of expertise, etc.(1). There has been a significant increase in pharmaceutical industry involvement in NTD product development(1), which has been attributed to social and ethical concerns, efforts to improve reputation, and long-term strategic considerations(7).

Much of the increase in funding has been directed towards advancing particular products through late-stage trials. Most of the funding directed towards controlling NTDs is spent internally by large pharmaceutical companies(1), so it is likely that one of the main ways of partnering with them is through some kind of technology transfer of an invention to the company. Much of the partnering done by large pharmaceutical firms is now shifting to earlier stages of development(1,5,7). One of the biggest strengths of this type of partner is that it is one of a few types of organisation that can advance a project from any point in the product life cycle to the market. The 20 or so large, multi-national corporations also have wide networks to engage CROs and other partners to carry out different parts of the development work. Therefore, these firms can be approached for partnership at any stage in development. Of the largest companies, work in the NTD space is dominated by four of them: GlaxoSmithKline, Novartis, Sanofi and Astra Zeneca. Although they have large amounts of money at their disposal, this means that there are really only a small number of individual developers of this type to partner with, compared with the biotechnology cohort.

Biotechnology/Small Pharmaceutical Companies

It has been shown that technology transfer from Universities to biotechnology firms accounts for more new drugs and biologics than that from Universities to pharmaceutical companies (16% vs. 8%) (34). Therefore, biotechnology companies should be considered as one of the primary partners for researchers seeking to develop NTD products. It's often the case that these companies will later partner with or transfer technology to large pharmaceutical companies, but some biotechnology companies will advance their products all the way from discovery to market(34). This highlights the extent of some of their capabilities. Funding from biotech and small-pharmaceutical firms accounted for 10.6% of 2011 funding from industry that was reported in the G-FINDER report, and made up large/majority contributions in some disease areas, such as helminth diseases or trachoma(1).

Compared to the biotechnology sector as a whole, a very small proportion of companies are involved in NTD product development. However, participation by biotechnology companies in NTD product development is estimated by BVGH to be evident across 41% of the NTD product pipeline. Their participation is also present through all phases of the pharmaceutical product development lifecycle, and it exceeds the participation of pharmaceutical companies at all stages of development before phase III. It is possible that biotechnology companies are filling very important roles in different parts of the lifecycle, sometimes where large industry participation is absent(38).

When it is appreciated that a main obstacle in finding funding to develop NTD products is a lack of commercial potential, it is encouraging that biotechnology firms have played a more significant role than large pharmaceutical firms in discovering and developing orphan drugs (drugs developed for rare diseases; see discussion on rare diseases and orphan drugs below), where the commercial value is typically lower(34). This is analogous to the situation that has drawn some small companies to neglected diseases (this was the case for Zentaris in the development of miltefosine, for example) (7). In lower-income countries, where NTDs are present, this may be a further encouragement to engage in NTD product R&D. In 2011, 28.9% of funding reported by biotechnology and small pharmaceutical firms came from countries in lower-income countries. Firms from the emerging markets of India and Brazil are playing a particularly prominent role, and there have been significant increases in research directed towards treating kinetoplastid diseases in Brazilian firms(1).

Biotechnology and other smaller firms are limited by the funds available to them. As their funds originate from investments, as opposed to the revenues of large pharmaceutical companies, they have less flexibility, thus typically contribute less to NTD product development(38). Furthermore, it can be expected that these firms are also less able to offer in-kind contributions than their larger counterparts. PDPs often have to contribute more funding to biotechnology firms, than to pharmaceutical firms, making them a more expensive partner for PDP brokers to work with(29,38).

Contract Research Organisations

Whether conducting neglected disease R&D or not, CROs can be a useful partner for carrying out specific studies that researchers may lack the ability to do in their own facilities. CROs are reasonably active in the field of NTD product R&D and partner extensively with PDPs: the DNDi website lists 28 different CROs from companies from both developed and developing countries as partners(53). An analysis of traceable external PDP expenditure in 2007 shows that 20.9% went to CROs, larger than that which went to large pharmaceutical companies (18%) and biotechnology/small pharmaceutical companies (9.3%) (29).

There are some companies that are combining the role of a traditional pharmaceutical or biotechnology company, with contract services. Founded with money from equity investments,

the South Africa-based iThemba raises additional money through contract medicinal chemistry services and invests the profit into research for treatments of unmet medical needs in Africa. Using this business model, they have successfully generated a pipeline of leads and molecular scaffolds active against malarial parasites, TB and HIV(54). Another company, called Scynexis, has played an important role in developing a late-stage HAT compound(55). This is detailed in a case study, later in this chapter.

Venture Capital

When a researcher's commercialisation strategy is to form a new company, after gaining seed/start-up capital, larger investments are often sought from venture capitalists(39).

Venture Capital (VC) firms provide finance in exchange for some form of equity in the company, but individual deals can take many forms. VC firms usually make an investment with the assumption that there will be a profitable 'exit', at a later date. This will usually involve the company being purchased at a value that is higher than that invested in the company.

In addition to Private VC firms, Corporate Venture Capital (CVC), coming from large corporations, is another source of equity. There has been an increasing trend in venture capital investments to biotechnology/small pharmaceutical firms from large corporations, particularly large pharmaceutical companies. Several large pharmaceutical companies now have venture capital arms. Corporate venturing can be seen as another way that large firms can augment their pipelines and access new technologies. As some private VC firms have moved away from the biotechnology sector, Corporate VC has become increasingly important. Investments are reportedly on the increase, with a shift towards earlier stages of development, but these firms are often looking for companies that are already venture-backed(56).

The lack of financial incentive for making such investments for companies in the NTD space is probably a reason why venture capital has not been looked at as a significant source for funding in this area. However, this does not preclude it from playing a role. There will be a number of cases mentioned in this chapter, where funds from equity investments have eventually played a role in NTD drug development. Additionally, there is already evidence of a venture capital firm engaging in the fight to control neglected diseases. The Legatum group, a Dubai-based private investment firm, is a founder of the END fund, which is a global fund committed to supporting mass drug administration to eradicate NTDs in Africa(57).

Angel Investors

Angels, wealthy individuals that are usually serial entrepreneurs, make investments in a similar manner to venture capitalists, and are another source of funding for new businesses. Like the individual philanthropists, they are individuals who wish to invest in causes that they have some kind of interest in. They often seek to have some influence on the company, too(39). For

reasons already discussed, such an individual is likely to be a philanthropist in the area of NTD product R&D.

7.2.8 Details of partnerships

7.2.8.1 NTD product development in the 21st century

One of the most striking things of Pedrique et al.'s analysis of new products for NTDs (see products in Table 7.2.2), is that there are very few new innovative products – the majority of products are repurposed drugs (denoted NI, for new indication), new fixed dose combinations (FDCs) of existing drugs or new formulations (NF) of existing drugs(5). At first glance, there appears to be a promising number of new vaccines introduced in the period studied. However, these only cover two diseases, and two vaccines originate from the same invention. There are only four NCEs on the whole list(5). As a result, this gives us little information on how a product may be developed from the discovery phase, through to product registration. On the other hand, they do show evidence of successful product development through the more expensive, later stages of development. Therefore, there are several useful lessons to be learned from these successes.

Table 7.2.2 - New NTD products approved or recommended during 2000-11. Table reproduced from the work of Pedrique et al.(5), with an additional column on product development history (from media reports, WHO documents and scientific literature).

Generic Name	Marketed Name(s)	Indication(s)	Formulation	Product Type	Regulatory Body(s)	Country	Date of Approval	MA Holder	Comments on Development History
artemotil		malaria	injectable	NCE	MEB	Netherlands	2000	Artecef BV	Artecef BV was created by ACF Beheer S.A. for the purpose of developing Artemotil with the WHO. Patent held by the WHO Scientific Working Group on the Chemotherapy of Malaria.
atovaquone plus proguanil (paediatric)	Malarone Paediatric	malaria	tablet	NI	FDA	USA	2000	GlaxoSmithKline	
proguanil plus chloroquine	Avloclor	malaria	tablet	FDC	AFSSAPS	France	2001	AstraZeneca	
chloroquine plus primaquine		malaria	tablet	FDC	ANVISA	Brazil	2001	Industria quimica do Estado de Goias S/A - Iquedo, Brazil	
artesunate		malaria	injectable	NF	ANVISA	Brazil	2001	Silvestre Labs Quimica e Farmaceutica Ltda, Brazil	
artemether plus lumefantrine	Coartem	malaria	dispersible tablet	NF	Swissmedic	Switzerland	2008	Novartis Pharma Schweiz AG	Ciba-Geigy, later Novartis entered into a partnership with the Academy of Military Medical Sciences in China and several private Chinese entities to develop Coartem. Originally marketed as Riamet for developed-world travellers, but was produced non-profitably

									later, following an agreement with the WHO. Supported by the Global Fund.
artesunate plus amodiaquine (ASAQ)	Coarsucam	malaria	tablet	FDC	WHO PQ		2008	Sanofi-Aventis Group	Combination developed in parallel by Sanofi and the FACT (Fixed Dose Artesunate Combination Therapy) Consortium - a group headed by DNDi, and involving Farmanguinhos/Fiocruz and various other academic, clinical and CRO partners. They later teamed up to co-develop the product. Compound registration and manufacture was also initiated in Morocco, to build capacity and aid registration in endemic countries.
artesunate plus mefloquine (ASMQ)		malaria	tablet	FDC	ANVISA	Brazil	2008	Farmanguinhos/ Fiocruz, Brazil	Combination devised by WHO TDR staff, with support from USAID and the Wellcome Trust. Developed in a PDP led by the FACT consortium. A later technology transfer agreement allowed production and registration in India.
artemether plus lumefantrine		malaria	oral suspension	NF	CDSCO	India	2008	Information unavailable	
arterolane maleate plus piperaquine phosphate	Synriam	malaria	tablet	NCE	CDSCO	India	2011	Ranbaxy labs	Invented in a PDP, led by MMV, with UK, US and Swiss contributors. MMV licensed the IP to Ranbaxy in 2003.

piperaquine tetraphosphate plus alpha dihydroartemisinin	Eurartesim	malaria	tablet	NCE	EMEA	Europe	2011	Sigma-Tau Industrie Farmaceutiche Riunite SpA	Combination initially developed by Professor Li Guoqiao of Guangzhou University of Traditional Chinese Medicine and developed in a partnership with the Holley group, China. Further developed for Global use in a PDP involving MMV.
artesunate		malaria	suppository	NF	ANVISA	Brazil	2000	Nova Quimica Farmaceutica Ltda, Brazil	Supporting research was funded by WHO TDR, several governments and two PSNRIs.
ethambutol plus isoniazid plus pyrazinamide plus rifampicin		tuberculosis	tablet	FDC	WHO PQ		2003	Lupin Ltd, India	Combination was introduced on advice of WHO Advisory committee.
ethambutol plus isoniazid plus rifampicin		tuberculosis	tablet	FDC	WHO PQ		2008	Macleods Pharmaceuticals Ltd, India	Application submitted by MSF and The International Union against Tuberculosis and Lung Disease.
isoniazid plus rifampicin		tuberculosis	dispersible tablet	NF	WHO PQ		2009	Macleods Pharmaceuticals Ltd, India	
isoniazid plus pyrazinamide plus rifampicin		tuberculosis	dispersible tablet	NF	WHO PQ		2009	Macleods Pharmaceuticals Ltd, India	
moxifloxacin		tuberculosis	tablet	NI	WHO PQ		2010	Cipla Ltd, India	Use is currently off-label, although recommended by WHO; Bayer (after withdrawing its patent in India) and TB alliance are conducting trials to register the product.
levofloxacin		tuberculosis	tablet	NI	WHO PQ		2011	Cipla Ltd, India	
ofloxacin		tuberculosis	tablet	NI	WHO PQ		2011	Cipla Ltd, India	

benznidazole (paediatric)		Chagas disease	dispersible tablet	NF	ANVISA	Brazil	2011	Pernambuco State Pharmaceutical Laboratory (LAFEPE), Brazil	Partnership with DNDi and multiple funders to produce new formulation. DNDi supported product registration.
miltefosine	Impavido	visceral leishmaniasis	capsule	NI	CDSCO	India	2002	German Remedies Ltd, Mumbai	WHO TDR worked closely with Zentaris AG (original inventor) to clinically develop miltefosine for visceral leishmaniasis. WHO TDR set up a development team and several Indian principal investigators were involved. Strong agreements on patient access were put in place. Zentaris licensed it to German Remedies for manufacture and Distribution.
paromomycin		visceral leishmaniasis	injectable	NI	CDSCO	India	2006	Gland Pharma United and One World Health (San Francisco [IOWH]) and the Bill & Melinda Gates Foundation	
miltefosine	Impavido	cutaneous leishmaniasis		NI	CDSCO	India	2008	German Remedies Ltd, Mumbai	Zentaris appear to have sponsored further trials in this indication independently of TDR.
nitazoxanide	Alinia	cryptosporidiosis/ giardiasis	oral suspension	NF	FDA	USA	2002	Romark Laboratories, LC	See below
nitazoxanide	Alinia	cryptosporidiosis/ giardiasis	tablet	NCE	FDA	USA	2004	Romark Laboratories, LC	Originally discovered by Jean-Francois Rossignol in the Pasteur Institute.

									Rossignol co-founded a Romarck to commercialise nitazoxanide.
tosufloxacin tosilate hydrate	Ozex	cholera	granules	NF	PMDA	Japan	2009	Toyama Chemical Co, Ltd	A reformulation of Ozex tablets.
bivalent inactivated vaccine, killed whole cells of *Vibrio cholerae* O1 and O139	Shanchol	cholera	oral suspension	vaccine	CDSCO	India	2009	Shantha Biotech	International Vaccine Institute altered mORCVAX formulation of Vabiotech to meet International standards, with backing from the Gates Foundation and various international development agencies. The technology was then transferred to Shantha for production.
live pentavalent vaccine, for prevention of G1, G2, G3, G4, and G-serotypes containing P1A[8], in Vero cells	RotaTeq	rotavirus	oral suspension	vaccine	EMEA	Europe	2006	Sanofi Pasteur MSD, SNC	Discovered by the Philadelphia Childrens Hospital and the Wistar Institute. Product licensed to Merck and Royalty Interest of Philadelphia Children's Hospital and part royalty interest of Wistar institute was later sold to 3rd parties.
Live, attenuated vaccine, rotavirus RIX4414 strain, for prevention of G1 and non-G1 serotypes (G3, G4, and G9), in Vero cells	Rotarix	rotavirus	oral suspension	vaccine	EMEA	Europe	2006	GlaxoSmithKline Biologicals SA	Licensed from Avant to GSK. This vaccine was first registered for use in Mexico and the Dominican Republic in 2004, and has been approved in more than 35 countries and the European Union.

Japanese encephalitis virus vaccine, freeze-dried, inactivated, Beijing-1 strain, Vero cell-derived	ENCEVAC	Japanese encephalitis	injectable	vaccine	PMDA	Japan	2006	Kaketsuken	
Japanese encephalitis virus vaccine, live chimeric viral vector	Imojev	Japanese encephalitis	injectable	vaccine	TGA	Australia	2008	Sanofi Pasteur Pty Ltd	Developed by Acambis as ChimeriVax-JE. Marketing and distribution license granted to Sanofi. Target markets in South-East Asia, presumably for-profit.
Japanese encephalitis virus vaccine, purified, formalin-inactivated, whole virus vaccine strain SA14-14-2, in Vero cells	Ixiaro	Japanese encephalitis	injectable	vaccine	EMEA	Europe	2009	Intercell AG	Intercell licensed relevant IP from Vaccgen International LLC and Sanofi Pasteur. Distribution by Novartis.
Japanese encephalitis virus vaccine, live, attenuated strain SA14-14-2, in primary hamster kidney cell cultures	mORCVAX	Japanese encephalitis	injectable	vaccine	CDSCO	India	2010	Curevin Pharma Pvt	Curevin are providing Vabiotech's vaccine (see Shanchol, above).
Crotalidae polyvalent immune fab (ovine)		snakebite	injectable	biologic (anti-venom)	FDA	USA	2000	Protherics Inc.	Developed for North American market.
nifurtimox (combination therapy with eflornithine		human African trypanosomiasis	tablet	NI	WHO EML		2009	Drugs for Neglected Diseases Initiative (DNDi)	Combination supported by research conducted by MSF and epicentre.
zinc		diarrhoea	tablet and liquid	NI	WHO EML		2006	Newborn and Child	Research performed by academics, supported by

						Development Department of Child and Adolescent Health, WHO	NGOs.
ribavirin	viral haemorrhagic fevers	intravenous and oral	NI	WHO EML	2007	Biorisk Reduction for Dangerous Pathogens Team (BDP), WHO	

Regarding the products listed in Table 7.2.2, it can be seen that a variety of different organisations filed for registration and hold marketing authorisations (MA) (i.e. are authorised to market and sell the product). The majority is made up of pharmaceutical manufacturers, mainly large pharmaceutical corporations and smaller firms in the industry. There are also some PDP brokers and transnational organisations, serving a similar role to PDP brokers, in the list: DNDi, the Newborn and Child Development Department of Child and Adolescent Health of the WHO, and the Biorisk Reduction for Dangerous Pathogens Team of the WHO.

The aminoglycoside antibiotic, paromomycin was launched in 2006 as a treatment for visceral leishmaniasis (VL) with joint marketing authorisation going to Gland Pharma United, the Institute for One World Health (IOWH) and the Gates Foundation. IOWH has a role as a PDP broker, but also carries out R&D activities. Originally discovered in the 1950s, paromomycin was shown to have efficacy in the treatment of VL in the early 1960s. There was a long history of use and development of paromomycin by trans-national NGOs and public health authorities, MSF and WHO TDR, but the discontinuation of paromomycin production and the cessation of development almost prevented this treatment from having a place in VL therapy. In 1999, IOWH began completion of paromomycin development and received funding from the Gates Foundation to realise this goal. They partnered with Indian pharmaceutical company, Gland Pharma, to ensure production(58,59). This is an example of a successful PDP, which highlights a number of key strategies. IOWH is a non-profit developer, so they were able to develop a treatment where a significant commercial incentive did not exist. They were able to do this, only with philanthropic support. It was also important to engage a producer in a country where VL was endemic.

The case of paromomycin highlights another important pattern seen in Table 7.2.2: 18 of the 33 industrial MA holders are based in developing countries where NTDs are prevalent, India and Brazil. These two countries are also the two largest developing country funders of NTD research(1).

Most of the products in Table 7.2.2 were developed in a partnership, or there was some form of technology transfer between different entities. Another trend seen here is that almost all products that are marketed in lower- or middle-income country markets and/or on a not-for-profit basis received some form of public, philanthropic or NGO support in their development. The WHO or PDPs were involved in supporting research (see artemotil, Synriam®, ASMQ, miltefosine and others) and product registrations (see benznidazole [paediatric]). State-owned pharmaceutical firms were involved in the development of several products in Brazil, and academic/PSNRIs were the home institutions of the original inventors of several of the products (e.g. artemisinin derivatives, such as artemether and artesunate). Even in some cases, where a product was originally targeted towards developed country markets (e.g. Coartem® for travellers), engagement by one of the aforementioned organisations (the WHO in the case of Coartem®) was the key factor in enabling patient access in lower-income countries.

The results of the upsurge in NTD drug development and the emergence of new strategies have not yet been seen in full. The analysis above provides a somewhat misleading picture of the current state of NTD drug development. The BVGH global health primers shows pharmaceutical development projects at every stage in the development life cycle. In March 2012, 190 drugs and 225 drugs were reported as being in 'active development' for 23 different NTDs(60). Thus it is important to analyse some of the partnerships of drugs in development. Fexinidazole, a new treatment for HAT is in late clinical trials and is a great example of the effectiveness of the PDP strategy for developing NTD products. Case studies of other products, which have required the use of innovative business models or strategies, are discussed near the end of this chapter.

Fexinidazole

Fexinidazole is being developed in a PDP driven by DNDi in partnership with Sanofi, the Swiss Tropical Health Institute and HAT platform partners, a consortium of at least 18 national HAT control programs, research institutes and NGOs(3,61). Fexinidazole, a broad spectrum nitroimidazole antibiotic, was in preclinical development to treat bacterial infections in the 1970s, before efficacy against *T. brucei* infection was discovered in the 1980s. It was not developed further for either indication, but it was re-discovered during a large characterisation of 700+ compounds, initiated by DNDi in 2005(62,63). DNDi was responsible for much of the R&D: preclinical, clinical and pharmaceutical. Sanofi carried out industrial development, registration and manufacturing. Members of the HAT platform, particularly the national control programs of the Democratic Republic of the Congo (DRC) and the Central African Republic (CAR), along with Médecins Sans Frontières (MSF) were responsible for clinical trials. DNDi list a number of funders, ranging from the Gates Foundation to government international aid departments(61). Fexinidazole is expected to be available on the market in 2014(64).

7.2.9 Benefits of different options

The presence of different potential partners at different stages of the development lifecycle means that several different approaches can be taken to commercialisation of research in this area. Whether the IP protecting an invention is sold, licensed, or used as a foundation for forming a new company, partnering with other organisations is essential. This is evident when looking at the development histories of most drugs for NTDs(4,7,29) (see Table 7.2.2), and is a key characteristic of pharmaceutical development when the IP originates from academic/PSNRI labs(34).

Chesbrough & Schwartz describe how an effective product co-development partnership should look to leverage the capabilities of the respective partners. In order for such partnerships to be successful, the business objectives of each partner should be aligned(65). An example of business objectives for the commercialisation of the research presented in this thesis are:

- Gain evidence of the efficacy of compounds that are effective in killing trypanosomes

- Optimise leads to increase efficacy
- Prove *in vitro* and *in vivo* safety of new leads
- Prove in-human efficacy of new anti-trypanosomal drug (clinical trials)
- Register new treatment for HAT/Chagas Disease
- Large-scale manufacture and distribution of new drug for HAT/Chagas Disease

Perhaps, a reason for the success of the partnership models used in NTD product development is the alignment of these objectives. Different partners can contribute to projects in a complimentary manner, utilising capabilities that are required for the success of the project, but without perceiving any threats to their core business model. For example, a reason for the collaboration by large pharmaceutical companies in early stages of R&D (i.e. discovery) is that projects are pre-competitive at that stage – the outcomes do not bring the collaborators into competition with each other.

The main questions for academic/PSNRI researchers to address when trying to commercialise research that could lead to a new NTD product, revolve around where to find funding, who to partner with and at what stage(s).

7.2.9.1 The most appropriate option?

Table 7.2.3 - Market segmentation matrix, adapted from G Moore's 'Crossing The Chasm'(66). This tool is usually used to identify an initial market segment for a new product launch. In a similar manner, here it is used to identify the most appropriate type of organisation to partner with. A higher score indicates a more favourable option.

Key Criteria for Partner Selection	Segment 1 **Research Councils and Philanthropists**	Segment 2 **PDP Brokers**	Segment 3 **Large Pharmaceutical Companies**	Segment 4 **Biotechnology Companies**
Is the partner well-funded? Are they readily accessible?	4	3	5	3
Do they have a compelling reason to partner with us?	5	5	3	2
Can we, with the help of the partner, deliver a whole product to the market?	2	5	5	3
Is there no entrenched competition that could prevent us from getting a fair shot?	3	3	4	4
If we enter into partnership, will their references allow us to leverage entry into additional partnerships with others?	4	5	4	3
Total Score (5=high/certainly, 3=medium, 1=low/unsure)	18	21	21	15

Using Moore's market segment rating grid(66), it is possible to compare the relative strengths of potential partners (or 'market segments'). PDP brokers and large pharmaceutical firms appear as the best initial partners. In reality, partnership with one type of third party does not preclude a partnership with another. The analysis does highlight PDP brokers as the most appropriate partner for a number of reasons. They have access to funds, they have extensive networks with which to find partners and they have project management capabilities. Large pharmaceutical companies share many of these capabilities. However, PDP brokers have a business model that is designed specifically for the development of NTD products. A large pharmaceutical company may actually be limited by their business model, which requires profit,

so a third party may need to be engaged to support non-profitable efforts, as has been the case with some NTD products developed in partnership with them.

From an academic/PSNRI researcher's perspective, an important advantage of engaging with one of the two preferred partners is that it relieves them of the significant effort of securing funds for several different development activities and then constructing partnership agreements with numerous third parties. One concern in "handing off" projects, particularly when IP is sold or licensed without further participation from the inventors, is a loss of control over the progression of the product's development. Large pharmaceutical companies are sometimes forced to discontinue the development of a project for financial reasons(29). In this respect, either engaging a PDP or ensuring that partnership agreements involve continued participation would be a safer option.

In NTD product R&D, royalties are generally not a concern for academics/PSNRIs. However, this would be more of a concern if a new company was formed around the IP. For the institution as a whole, forming a new company and taking equity in it can have a significantly higher return than licensing to third parties(67). Given the desire for profitability for traditional start-up funders (i.e. venture capitalists) it would not be advisable to follow this strategy, in the absence of a product that can become profitable.

From a public health perspective, the most favourable outcome is an effective new drug to treat trypanosomiasis that is made available to the target population within the shortest time possible, and at the lowest cost to all parties involved. It is clear that the available evidence strongly points to the PDP model as the most effective approach to achieve this(7,29).

7.2.10 Evidence for Partnerships

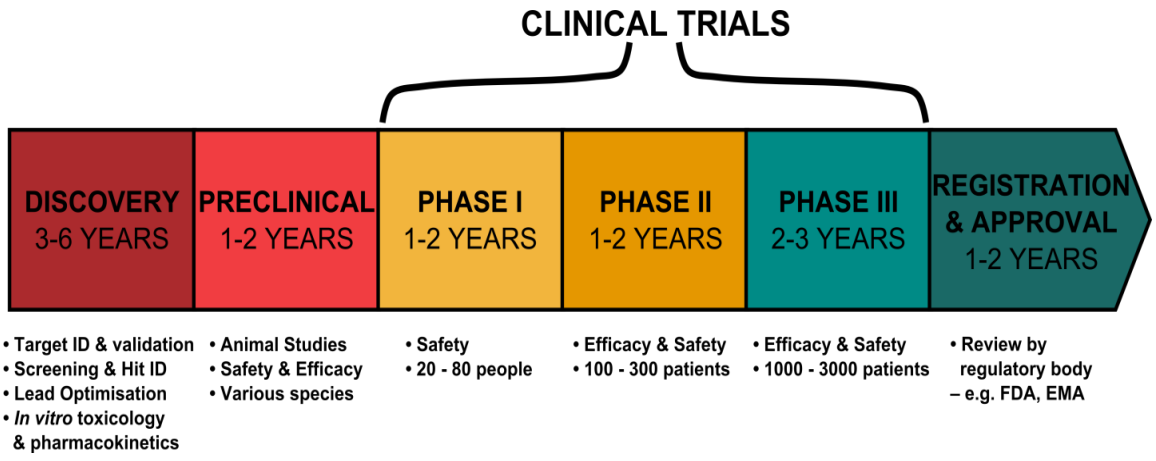


Figure 7.2.5 - Summary of the pharmaceutical development lifecycle, from discovery to registration and approval. Timelines and patient numbers informed by DiMasi et al.(68), S Blank(69) and the FDA(70).

7.2.10.1 Technology transfer in the pharmaceutical industry

Co-development partnerships are not unique to the NTD field, where pharmaceuticals are concerned. When seeking a partner, it is important to understand their motivations for entering

a partnership, as well as anything that may discourage a potential partner from entering into an agreement. An analysis of the evidence to support a particular project is warranted.

Pharmaceutical development can be an incredibly costly and risky process. An analysis of drug development success rates showed that only 19% of drugs entering clinical trials between 1993 and 2004 eventually attained clinical approval. This number was a little higher for co-developed (in-licensed or out-licensed drugs), than for drugs developed solely by one of the 50 pharmaceutical companies analysed in the study (27% vs. 16%). This may indicate that the licensors of new drugs usually prove efficacy at phase I or phase II, before licensing out to a third party. Encouragingly, systemic anti-infective drugs had a higher success rate of 24%, particularly at the transition from phase II to phase III development. This may be due to clearer clinical end points(71). Achieving the dual objectives of forming an effective partnership and eventually gaining clinical approval rely heavily on gaining evidence to support your project.

In an analysis of biologic drug licensing deals from biotechnology to large pharmaceutical companies, Kalamus and Pinkus report that the majority of deals occur between the preclinical phase and phase II. However, phase II appears to be the preferred time(72).

In their analysis, Kalamas and Pinkus show that both parties could generate value by entering into agreements earlier. Under-pricing by large pharmaceutical companies was stated as a cause of delayed deal making(72). With this problem removed in the case of not-for profit endeavours, deals should be made earlier, so that researchers can start to exploit the capabilities of their better-resourced partners sooner.

It has been reported that, after a shift towards deals made after in-human proof-of-concept, early-stage pharmaceutical industry deals are re-emerging: they represented more than 60% of 'big buyer' deals in 2011-2012. Large pharmaceutical companies are also forming more early-stage partnerships directly with academia. These partnerships often involve payments at milestones, when key research objectives have been met. They are sometimes tied to 'options' where, at the completion of all research objectives (i.e. phase II trials), the company can acquire the IP, or even an IP-holding company (where a small company is involved). This kind of arrangement allows the large companies to reduce their exposure to risk, whilst still enabling researchers to advance projects(73).

7.2.10.2 When to partner and with what supporting evidence

The most favourable time to start partnering largely depends on the commercialisation strategy followed by the researchers. In the absence of commercial considerations, it would even be advantageous to seek partners at the discovery phase. If an academic/PSNRI researcher is developing a product that may have some commercial value, then they may wish to attempt technology transfer later in the development lifecycle. In either case, researchers in an academic/PSNRI or in a small company may reach what is often termed "The Valley of Death" or the "Darwinian Sea", which is the point where more funds are required to develop a product,

but there is a lack of willing partners to fund the development of a product, which may still carry a high risk of failure(74,75). As highlighted by Kalamas & Pinkus, having the opportunity to use the capabilities of a well-resourced partner, early in the development process, should also be seen as an advantage(72).

The vast sums of money and significant amount of labour required to optimise products and demonstrate their safety can be out of the reach of most academic/PSNRI researchers(27). It can cost hundreds of thousands of dollars to outsource labour-intensive tasks, and *in vitro* and *in vivo* toxicology studies can cost over US$ 1million(36). Animal studies, which can be very expensive, can play a vital role in demonstrating efficacy(74). In these situations, an alternative to licensing to industry is to partner with a well-resourced academic drug discovery centre to carry out some of this work(27). There are also a number of programs to support translational research originating from academic/PSNRI research. One such program is the NIH's Therapeutics for Rare and Neglected Diseases program (TRND). TRND aims to collaborate with researchers of rare and neglected diseases "with the goal of moving promising therapeutics into human trials"(76).

Aside from gaining access to funding, there are other advantages to partnering with larger partners (e.g. pharmaceutical companies) which are due to the partner's greater experience. A review of strategic alliances between biotech companies and pharmaceutical companies between 1988 and 2000, gives some clues as to what these advantages are. It was shown that in the more complex trials of phase II and phase III, a drug is significantly more likely to proceed to the next phase, if the licensee (e.g. large pharmaceutical company) has more experience. In phase I trials, which are less complex, there is only a small advantage in this respect(77).

An advantage of being affiliated to academic institutions or PSNRI is that the capabilities (e.g. biological and clinical expertise) of the institution can be made use of at several points in the lifecycle, as has been done in the development of other products for NTDs or low-income markets (78). Many academic institutions are now engaged in translational research and there are several established 'drug discovery centres' at institutions across the US(27,28). However, once large scale trials are needed, these institutions become limited, so one should seek to partner with a PDP broker and/or large pharmaceutical company, and an organisation experienced in public health (e.g. WHO, MSF).

There are now several studies that can be used to 'de-risk' drug development projects(79). They are often used in a standardised way in the pharmaceutical industry, but it is often said that, in respect to these studies, expectations held by licensees in large pharmaceutical companies and their licensors, are mismatched. There is little specific information on these expectations and how they affect the partnering decisions of funders of NTD product R&D. Thus, this will be the subject of future research.

7.2.11 The role of other incentives

Clearly, a fundamental problem in funding R&D for NTD products is the lack of financial incentive for investors to support it. As discussed above, this limits the number of options a product developer can choose from to advance their product towards the market. In recognition of this long-standing problem, several different policies and market incentives have been proposed to create a market incentive for the development of products to meet an unmet medical need, where little/no market incentive exists(80–82). Amongst numerous proposals, two market incentives that have been implemented are the FDA's Priority Review Voucher (PRV) and Advance Purchase/Market Commitments (AMCs). They are both examples of "pull" mechanisms, which provide monetary incentives for the completion of research.

7.2.11.1 Priority Review Voucher

When the FDA considers a new drug to provide "significant improvements in the safety or effectiveness of the treatment, diagnosis, or prevention of serious conditions when compared to standard applications", it can be reviewed under 'priority review'. This means that the FDA promises to take action on a new drug application within 6 months of receiving it, compared to the standard 10 months(83). This can have the effect of reducing the overall time of review by 4 to 12 month, thus saving companies significant amounts of money(6). The PRV is a voucher granted to successful applicants of new drugs or vaccines that confer to the FDA's guidelines: they "offer major advances in treatment or…where no adequate therapy exists" for a specified list of tropical diseases. The PRV can be traded or sold, and allows the final user to gain priority review for a new drug application that wouldn't otherwise receive priority review designation. At present there is some uncertainty on the impact that the PRV will have. It has been estimated that the value of a PRV could range between US$50 and US$500 million, depending on the therapeutic area that the PRV is used in. We also lack information from past experience. Novartis were the first organisation to be awarded a PRV in 2009 following the approval of Coartem®, but they used that to gain priority review for another Novartis product. Given the potential for a high value PRV, the utility of this incentive can be substantial from an NTD researcher's perspective. However, the PRV is not awarded until development is complete, so it does not alleviate the shortage of funds during earlier- stage development(22). The PRV could be of some advantage to researchers using a spin-out company strategy, as its potential value could be used to attract the equity investments required to fund early-stage R&D. However there has been some criticism and scepticism shown towards the PRV program(21,22). It is not entirely certain how investors will respond to this program(81). Conversely, a survey conducted to assess the impact of the PRV program showed that the PRV had positively influenced decisions to continue investment in R&D for NTD products. The PRV was not a primary reason for pursuing NTD product R&D(22) and a separate study estimated that the PRV would not be sufficient to make an R&D venture profitable on its own(81). Nevertheless, the PRV could be a useful addition to a researcher's tools for attracting support for R&D.

7.2.11.2 Advance Purchase/Market Commitments

AMCs have been proposed as another way to increase private-sector participation in NTD R&D. An AMC is a legally-binding agreement that ensures that a company is paid a set amount of money for providing enough product to treat a set amount of patients, provided that the company produces a product that meets agreed specifications(18,19,80). This incentivises R&D by guaranteeing some return on a company's investment. Although AMCs remove market uncertainty for the investor, their payment still relies on a product being successful, and we have little evidence pertaining to how investors are responding to this. An AMC has proven to be successful in encouraging development of a Meningococcal C vaccine in the UK(80), and the first AMC for a neglected disease was implemented for pneumococcal vaccine development in 2007(84). Many of the potential advantages of PRVs apply for AMCs, but again, they do not supply up-front funds to finance R&D. They do have an additional advantage from a public health perspective, which is that they encourage rapid access for patients requiring the new product(19).

7.2.11.3 Research prizes

Research prizes or medical innovation prizes have been proposed as another mechanism to encourage research into treatments for diseases with unmet medical need. Furthermore, they have been seen as a way to widen patient access to innovative medicines by de-linking R&D costs from the prices of the medical products that they create(85–87). These prizes award a fixed sum of money to the individual or group who matches the specific research goal of the competition. There are numerous historical examples of prizes given for specific research outcomes that provide some benefit to human health(88). The setting up of research prize funds has been debated at the national and international level, but no large scale, public-sector system exists yet(85). However, there are prize funds or innovation prize-granting organisations operating in the private sector. Two of the largest of these are the Innovation Management firm, Innocentive(89), and the non-profit Xprize fund(90). These organisations advertise specific innovation problems, which are attached to prizes of a specific value. A researcher or company looking to advance a product already in development would need to hope that their research aims are aligned with those of a particular competition. Alternatively, researchers may choose to begin research in response to a competition. As of yet, no specific prize funds exist for NTDs, but the experience of the non-profit Prize4Life fund may prove the effectiveness of this approach. Prize4Life offers a prize of US$1 million for discoveries that advance the understanding and treatment of the debilitating and life-threatening disease, amyotrophic lateral sclerosis (ALS). A US$1 million dollar prize was awarded in 2011 for the discovery of an ALS biomarker. A prize fund of US$1 million was reopened in 2012 to find an ALS treatment, following the failure to find an effective treatment during the last competition, which occurred in 2008-10(91). Researchers should remain alert to competitions posted by prize-giving organisations, such as Innocentive, Nesta(92) and Xprize.

7.2.12 Roads less travelled

The very particular difficulties faced by researchers in the NTD space have demanded that a number of innovative commercialisation strategies are attempted. In some cases, a new company has been set up with an innovative business model that facilitates the development of a product with little/no profit-making potential. For example, in order to develop the antimalarial artemotil, ACF Beheer SA formed a new company, Artecef BV, which partnered with the WHO TDR program to bring the product to market. The following section gives further case studies and examples of the different strategies used by researchers to develop pharmaceuticals and medical devices for patient populations with little purchasing power.

7.2.12.1 The role of different indications/applications

The findings of the G-FINDER report suggest that, at least amongst industry funders of NTD product R&D, there is a shift towards more contributions to projects that might result in a semi-commercial (profitable) product. The analysis of the drugs studied by Pedrique and colleagues in Table 7.2.2 also highlights this trend towards the development of products with profitable applications. When a product has an additional indication for a disease where there is a market with commercial value, commercialisation for the profitable application can help to fund the development or production of that product for its original intended use. A factor contributing to Sanofi making eflornithine freely available to HAT patients, was that they were producing the drug as part of their Vaniqa® product, which is used to treat hirsutism in women (primarily in developed countries)(93). Other products, such as several of the malaria treatments on the market, find application in the developed world, which can help subsidise the donation of those same products to developing-country markets. Below, creative commercialisation strategies which exploit these situations are described.

Case Study: The Infectious Disease Research Institute (IDRI)

The IDRI is a non-profit research institute that specialises in applying knowledge of immunology to the discovery and development of products to diagnose, prevent and treat neglected tropical diseases, including Chagas Disease and Leishmaniasis. They benefited from their non-profit status, by being able to engage with stakeholders involved in the use of their products, but they suffered from the inconsistency of the grant funding that was available to them. To bring in larger sums and more sustainable sources of money, they span out companies to commercialise products that had profitable developed-world applications. IDRI licensed their vaccine adjuvant technologies to these companies and sometimes held equity in them, so the money earned from royalties or acquisitions could be used to serve their NTD product development aims(94).

A trypanosomiasis treatment with additional indications

A product resulting from the research presented in this thesis could potentially have uses for other disease indications. The results of BLAST searches using the L-threonine 3-dehydrogenase (TDH) protein or gene sequences from *Trypanosoma brucei*, show that similar proteins and genes have been identified in *Staphylococcus aureus, Clostridium difficile, Enterococcus faecium* and *Enterococcus faecialis*. These organisms are all associated with nosocomial infections and represent urgent medical needs. There is a strong case for new antibiotics to treat infections by drug resistant bacteria, such as methicillin-resistant *Staph aureus* (MRSA) and vancomycin-resistant *Enterococcus* (VRE)(95,96).

The global market for therapeutics for MRSA infection has been estimated to grow from US$2659.7 million in 2011 to US$3468 million in 2019(97) and the global market for therapeutics for *C. difficile* infections has been set to grow from US$314.2 million to US$314.2 million, over the same period(98). If a treatment developed for trypanosomiasis came from this research, and it was also effective in killing these bacterial pathogens, there would be clear commercial potential. In this case, a sensible strategy would be to license the IP to for-profit spin outs, as in the IDRI case study. If they were successful in attracting investment, the money gained through licensing could be used to further the development of the trypanosomiasis treatment. Spin outs would also allow the University to recover more value from the commercialisation of these products. Similar scenarios can also be envisaged for drugs which targeted other targets shared amongst several important pathogens.

7.2.12.2 Rare disease/orphan drug strategy

The Orphan Drug Legislation passed in the US and the EU established packages of push and pull mechanisms to spur drug development for rare diseases, where the market size is too small to expect a significant return on investment. Drugs for these rare diseases are termed ‘Orphan Drugs’. The criteria and benefits of each piece of legislation are outlined in Table 7.2.4. Several other countries have enacted similar laws to achieve the same goals. These laws have been successful in encouraging the development of orphan drugs, which may not have been developed otherwise, due to a lack of commercial viability. Pursuing research in this area has also been adopted as a commercial strategy for several biotechnology and pharmaceutical firms, for whom orphan drugs may be as profitable as non-orphan drugs(28,99–101). Due to their low incidence in developed countries, several NTDs are classified under the aforementioned legislation as rare diseases; several NTD drug and vaccine development programs have taken advantage of some of the benefits provided under the legislation(80). As with orphan drugs, regulatory authorities may also apply some flexibility when approving products for NTD indications(102). However, orphan drug legislation has been nowhere near as effective in promoting R&D for NTDs, as it has for other rare diseases. In developed countries, orphan drug developers have taken advantage of market exclusivity and the purchasing power of public health authorities or insurance companies to charge extremely high

prices(103). This sort of financial incentive does not exist for products for NTDs, thus the utility of the orphan drug legislation is limited, and the benefits should be exploited alongside other mechanisms(104). The push mechanisms tied to orphan drug policies are particularly useful to NTD product developers as they offset several of the product development costs.

Table 7.2.4. Summaries of the designation criteria and incentives of the USA and EU Orphan Drug/Rare Disease laws(105–107).

	FDA – USA	EMA – European Union
Designation Criteria	• Disease affects fewer than 200 000 people in the US, or • Product is a vaccine, diagnostic or preventative drug and is intended to be administered to less than 200 000 people per year in the US, or • Disease affects more than 200 000 people or product is a preventative drug intended to treat more than 200 000 people per year, and R&D costs cannot be expected to be recovered by sales in the US.	• Disease must be life-threatening or chronically debilitating, and • Prevalence of the disease must be less than 5 in 10 000 in the EU, or it must be likely that marketing would not generate significant returns on the investment for development, and • No satisfactory method of diagnosis, prevention or treatment is available, or, if such a method exists, the medicine must be of significant benefit to those affected by the condition.
Benefits/Incentives	• 50% tax credit for costs of clinical trials in the US • 7-year market exclusivity – protection from competition from a drug for the same indication • Fast-track review of registration application • Written recommendations given on the preclinical studies and clinical trials required.	• Protocol assistance • 10-year market exclusivity – protection from competition from similar drugs with similar indications • 2-year extension of market exclusivity if paediatric investigation plan followed • More incentives for SMEs – administrative and procedural assistance. • Some EU member states (including the UK), offer tax credits for R&D

Case Study: Niprisan

Although not an infectious disease or NTD, sickle cell anaemia is a disease with a high prevalence in several lower-income countries in Africa and suffers from a shortage of safe and effective treatments. In Nigeria, where 150 000 children are born with sickle cell anaemia annually, the country's National Institute for Pharmaceutical Research and Development (NIPRD) identified a traditional herbal treatment that was effective in treating the disease. In collaboration with the Philadelphia Children's Hospital, NIPRD obtained evidence of *in vivo* efficacy for the treatment in animal models and later, in humans. The treatment, named Niprisan, was licensed to US biotechnology firm, Xechem, which gained orphan drug designation for Niprisan in the US, due to the low prevalence of sickle cell anaemia there, and developed it to the point of gaining FDA approval. Unfortunately, these factors

were not enough to ensure success. The product did not seem to be economically viable and Xechem filed for bankruptcy some years later. However, this is an example of how orphan drug legislation can be used to aid drug development. The reported benefits of the designation were: "waiving of registration fees, additional funding, and an increase in investor confidence"(108).

7.2.12.3 Multiple Indications

There is another way in which an Orphan Drug Designation may be useful, from a commercial standpoint. The development of a drug for an NTD or rare disease can be accelerated and partly funded through receipt of an orphan drug designation, then that same drug may be developed for use in a more common disease (e.g. bacterial infection). Alternatively, the knowledge gained through development of the original drug can be applied in the pursuit of drugs for more common diseases. This strategy is currently being applied with rare diseases by some companies in the pharmaceutical industry(101). It is therefore plausible that an NTD drug with other indications could help to attract an industry partner.

7.2.12.4 Exploiting technologies

Case study: Anacor Pharmaceuticals

The case of Anacor Pharmaceuticals, a US-based biotechnology firm, provides an example of how a for-profit company is able to leverage its technology to develop treatments for NTDs, on a not-for-profit basis. Anacor was set up around a boron chemistry platform that was discovered through academic research. The unique, boron-based compounds that Anacor holds the IP for have a distinctive geometry and reactivity, and have demonstrated efficacy in several diseases. To allow Anacor to apply this technology to NTDs, it entered into an agreement with DNDi, to whom it offered a non-exclusive, royalty-free license. This class of the compounds, termed oxaboroles, were found to hold much potential as HAT treatments. DNDi partnered with the CRO, Scynexis to develop this further and the resulting lead, SCYX-7158 has entered in-human trials(55). For Anacor to be able to increase their participation in the NTD space, the leadership constructed a business plan, where grants from PDPs and other NTD R&D funders would allow them to carry out NTD R&D without interfering with their shareholder-funded commercial R&D activities(109).

This example highlights a number of key points. Principally, it underlines the role of PDPs and other funding sources in engaging for-profit entities in the industrial sector. It also shows how alteration of a for-profit firm's business model can allow their participation in NTD product R&D. Other similar examples of business model alignment can be found in the Global Health Innovation Insight Series, produced by the Centre for Social Innovation, at Stanford University's Business School(110).

Equity investments by the Gates Foundation

In 2011, the Gates Foundation made its first equity investment in a biotechnology firm. Liquidia, a company based in North Carolina, USA, have attracted interest from several investors, for its novel vaccine technology. The purpose of the US$10 million investment by the Gates Foundation is to support the application of Liquidia's technology to diseases where there is no viable market(111). According to foundation executives, this is the first case of a new approach to supporting product development for global health(112) and this may represent a new source of equity for a for-profit company engaged in NTD product development. In 2013, the foundation bought US$5 million of common stock in Anacor Pharmaceuticals as part of an agreement that included additional grants of US$17.7 million to support NTD drug discovery and compound library development(113).

7.2.12.5 Open Source Drug Discovery

Open Source Drug Discovery is being used as a way to foster collaboration and to take advantage of the different skills of researchers involved in drug discovery. The six laws of Open Source Drug Discovery are:

- All data are open and all ideas are shared
- Anyone can take part at any level of the project
- There will be no patents
- Suggestions are the best forms of criticism
- Public discussion is much more valuable than private email
- The project is bigger than, and is not owned by, any given laboratory.

Although there are several concerns over whether the lack of intellectual property will make it difficult to attract funders, development partners and manufacturers, the open source approach has been successful in achieving the aims of combining the abilities and resources of a wide range of researchers(114). This has found utility, particularly in NTD drug development. Perhaps the most well-known open-source drug discovery project is the Open Source Malaria consortium. This project is funded by MMV, centred in Sydney and involves contributors from around the world(114,115). This campaign has resulted in the chemical synthesis of various leads, some of which have been tested *in vivo* (animal models)(116). A similar project, looking for new drugs to treat tuberculosis is also active(117). The Tropical Disease Initiative is a further example of an open source initiative with similar aims, targeted at a broader set of diseases(118,119).

As all of these examples are in the early stages of development, it is difficult to see how successful they will be in delivering new treatments to the clinic. Setting up an interface for a whole community of researchers can be an endeavour requiring significant effort. However, the Open Source approach can be a potentially useful option for a researcher who is limited by the capabilities available to them.

7.2.12.6 Crowd-funding

Crowd-funding, the process where large numbers of investors make small investments in a particular venture, as opposed to a few investors contributing large amounts, is a growing trend that is beginning to gain traction in the biotechnology and pharmaceutical industries. It has been proposed as another source of funding for early-stage product development(120). In 2012, 200 investors

gained a profitable exit through a French organisation, called WiSEED when an angel investor bought their equity stake in antibacterial drug discovery company, ANTABIO(121). There are now proposals to set up tax-deducted crowd-funding platforms in the UK: Citizen's Innovation Funds(122). For investors to buy equity in a private limited company, the law must allow for this. Legislation that does allow private companies to raise money through crowd-funding has been passed in the US(121).

Crowd-funding does not always involve equity investments, so it could be used by researchers to support specific R&D activities. In recent years, two web-based crowd-funding platforms, Petridish.org and Experiment.com have been established to allow scientists to crowd-fund individual research projects. Founded in response to funding shortages and a lack of support for certain ideas, Experiment.com proves that this approach can be successfully utilised to fund NTD product R&D. One researcher sought US$6500 to fund the testing of an FDA-approved drug in an animal model of Chagas Disease. Her success in raising the funds allows her to clear the final "obstacle…before the drug can be used in clinical trials"(123). Although the amount raised is relatively small, this is a clear example of how crowd-funding may be used to bring one's research to a point where it is easier to seek other partners/funders. The website has also been used to successfully raise money for projects targeting malaria and helminth infections(124).

7.2.13 Lessons from other industries

Evidence of the different strategies for advancing development of NTD products has largely been drawn from experience within this area and in pharmaceutical R&D in general. However, there are other lessons to be learned from the experiences of businesses in areas that are not specific to pharmaceutical R&D. Additionally, strategies from charitable/social sector organisations are also described below.

7.2.13.1 Support for start-ups and small businesses

When seeking seed capital and/or funds for early-stage product development, there are a number of sources of support in the public and private sector. There are several government-funded initiatives to promote the foundation of new companies, particularly in technology. For example, the UK government's Technology Strategy Board offers R&D funding in a number of strategic areas to academics and small businesses(125).

Business incubators aid the development of early-stage companies by giving various types of support, ranging from expertise and facilities to funding and mentoring. Incubators are usually run by government, universities, private consortiums or large corporations(126,127). They now play a prominent role in the development of technology start-ups and are increasingly being used by large pharmaceutical companies to partner with smaller firms and bulk up their pipelines(128).

Business Plan Competitions are another source of important seed funds or R&D funds. A useful example of this can be seen in the founding of Respira Design, a non-profit company specialising in

the design of low-cost devices for asthma treatments for use in low-income settings. When developing their first product, the founders struggled to gain support from investors to fund proof-of-concept studies. They were able to find this money by reaching the finals of the Ashoka Changemakers competition(78).

7.2.13.2 University alumni donations

Alumni donations provide a substantial source of income for universities and PSNRIs. They are often used to fund specific initiatives and may be tapped into to support NTD programs. The University of British Columbia's Neglected Global Diseases Initiative, a cross-departmental research group, was built through donations from alumni and other donors(129).

7.2.13.3 Microdonations

Just as Crowd-funding aims to raise relatively large amounts of money, from relatively small contributions and from a large number of investors/donors, microdonations raise money using a similar philosophy. A microdonation is a "low value donation 'embedded' into activities or transactions that consumers are already undertaking". The Pennies fundraising scheme, run by UK-based The Pennies Foundation, raised more than £250 000 in 12 months by asking consumers to donate a penny when making purchases in the UK(130). Similar initiatives have been used to raise funds to fight HIV/AIDS and tuberculosis(131,132).

The consumer activity at university-owned or university-housed stores could be harnessed in a similar manner to raise money for research. This would require a significant organisational effort, however, but it could raise substantial funds.

7.2.13.4 Social Investment

The utility of the start-up strategy for developing a new treatment is limited by the lack of profitability of the target market for trypanosomiasis. However, a non-profit business model could be useful if this strategy were pursued. Like academic/PSNRI researchers, charities and social enterprises are in constant search of funds to carry out their work. Furthermore they need particularly large amounts of money in order to expand their capabilities. These organisations are now being aided by a recent trend termed 'Social Investment'. Funds are given in the form of long-term debt, which is similar to an equity investment(133). Many of these investments require that the recipient has an underlying source of revenue, therefore excluding a pre-revenue pharmaceutical/biotechnology start-up. However, a new type of social investment, the Social Impact Bond (SIB), is becoming a useful tool for some non-profits. SIBs, which were founded by a group called Social Finance in the UK, are issued by the public sector to private sector investors, for work carried out by charitable organisations. The original investors only receive a return if the work of the charity has a measurable social impact that is cost-saving for the government. In this way, the charity receives up-front funds but bears no financial risk(134,135). As this is a relatively new initiative, it does not seem to have received any attention from the drug development or NTD communities, but it could feasibly be used for financing R&D.

7.2.14 Conclusion

The NTD R&D landscape

A dramatic increase in awareness and advocacy for NTD product R&D has been witnessed over the last decade. This is evident in the increased activity in the area and the evolution in product development and commercialisation strategies. Public-Private partnerships have always been a feature of NTD product R&D, but the emergence of the PDP as the primary model for product development is welcome progress. Unfortunately, reports of funding inconsistencies and research gaps for some diseases persist(5,38). Following the global financial crisis, there has been a decrease in funding from philanthropic sources and some government aid agencies. This is of concern because they are the principal funders of PDPs(1). Despite these concerns, researchers working towards developing products for NTDs are in a better position to do so than a decade ago. In addition to the increased funding and awareness, the experience of others is providing evidence of several new ways to attract funding to the cause. Researchers and other stakeholders of NTD product development should also look to the future, as new incentives, such as tax credits(136,137) and Medical R&D treaties(138), could result from recent proposals.

Limitations of this analysis

This chapter has aimed at providing insight into the main considerations to be taken account of when developing a treatment for a NTD. Thus, many aspects of pharmaceutical R&D, and how it is carried out for NTDs, have been broadly reviewed. However, in reality, some concerns, such as the specifics of licensing deals, intellectual property and target product profiles all demand deep analyses of their own. These concerns have all been eluded to here, but deeper analyses have focused on the most important and immediate problems faced when advancing this research. Specific evidence on the best stage of development to transfer technology was found to be lacking, so this will be the subject of future research.

7.2.14.1 How to commercialise this research

Opinion on the best course of action

Given the therapeutic area, the most appropriate mechanism for commercialising the results of this research would be through licensing to external partners. In the absence of other, profitable indications, the effort and seed capital required to found a spin-out company would not be worth it. If the drug/series of compounds showed efficacy in diseases with a market, then commercialisation along those lines should be carried out separately through licensing of the original IP. Profits from those endeavours could be used to support development of the original product.

A large, well-resourced partner, such as a PDP broker or large pharmaceutical company, may be sought for partnership as early as possible to make use of their capabilities and accelerate development. Ownership of IP and research objectives should be agreed upon and aligned as early as possible, to prevent any issues once co-development starts.

PDP brokers seem to engage with researchers once a viable product (e.g. a lead compound with some *in vitro* efficacy) has been found. Therefore, with this research, a partnership with a pharmaceutical or biotechnology firm should be used to bring the research to this point.

The greatest possible number of funding sources, government incentives and other support should be taken advantage of. The options will be different, depending on the commercialisation route followed; different options are available to non-profit organisations than to for-profit organisations. Drawing support from numerous sources is advantageous and may prove to be essential to advance a product all the way to the market.

7.2.15 References

1. George Institute. Global Funding of Innovation for Neglected Diseases (G-FINDER) [Internet]. The George Institute for Global Health. 2013 [cited 2013 Nov 30]. Available from: http://www.georgeinstitute.org/projects/global-funding-of-innovation-for-neglected-diseases-g-finder
2. Institute G. Global Funding of Innovation for Neglected Diseases (G-FINDER) database open to public [Internet]. The George Institute for Global Health. 2010 [cited 2014 Jan 15]. Available from: http://www.georgeinstitute.org/media-releases/global-funding-of-innovation-for-neglected-diseases-g-finder-database-open-to-public
3. BIO Ventures for Global Health. Global Health Primer [Internet]. [cited 2014 Jan 15]. Available from: http://www.bvgh.org/Biopharmaceutical-Solutions/Global-Health-Primer.aspx
4. Trouiller P, Olliaro P, Torreele E, Orbinski J, Laing R, Ford N. Drug development for neglected diseases: a deficient market and a public-health policy failure. The Lancet. 2002 Jun 22;359(9324):2188–94.
5. Pedrique B, Strub-Wourgaft N, Some C, Olliaro P, Trouiller P, Ford N, et al. The drug and vaccine landscape for neglected diseases (2000–11): a systematic assessment. Lancet Glob Health [Internet]. 2013 Oct [cited 2013 Nov 12]; Available from: http://linkinghub.elsevier.com/retrieve/pii/S2214109X13700780
6. Stefanakis R, Robertson AS, Ponder EL, Moree M. Analysis of Neglected Tropical Disease Drug and Vaccine Development Pipelines to Predict Issuance of FDA Priority Review Vouchers over the Next Decade. PLoS Negl Trop Dis. 2012 Oct 25;6(10):e1803.
7. Moran M. A Breakthrough in R&D for Neglected Diseases: New Ways to Get the Drugs We Need. PLoS Med. 2005 Sep 8;2(9):e302.
8. WHO | The 17 neglected tropical diseases [Internet]. WHO. [cited 2013 Dec 10]. Available from: http://www.who.int/neglected_diseases/diseases/en/
9. Hotez PJ, Fenwick A, Savioli L, Molyneux DH. Rescuing the bottom billion through control of neglected tropical diseases. The Lancet. 2009 May 2;373(9674):1570–5.
10. PLOS Neglected Tropical Diseases: A Peer-Reviewed Open-Access Journal [Internet]. [cited 2014 Jan 15]. Available from: http://www.plosntds.org/static/scope

11. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2012 Dec 15;380(9859):2197–223.
12. WHO | Chagas disease (American trypanosomiasis) [Internet]. [cited 2010 Dec 2]. Available from: http://www.who.int/mediacentre/factsheets/fs340/en/index.html
13. Simarro PP, Jannin J, Cattand P. Eliminating Human African Trypanosomiasis: Where Do We Stand and What Comes Next? PLoS Med. 2008 Feb 26;5(2):e55.
14. Adhoc Committee (of the WHO) on health research relating to future intervention options. Investing in health research and development [Internet]. Geneva: WHO; 1996 [cited 2014 Feb 4]. Available from: http://www.who.int/tdr/publications/tdr-research-publications/investing-in-health/en/
15. Fehr A, Thürmann P, Razum O. Expert Delphi survey on research and development into drugs for neglected diseases. BMC Health Serv Res. 2011 Nov 16;11(1):312.
16. Hotez PJ, Pecoul B. “Manifesto” for Advancing the Control and Elimination of Neglected Tropical Diseases. PLoS Negl Trop Dis. 2010 May 25;4(5).
17. The PLoS Medicine Editors. The Neglected Diseases Section in PLoS Medicine: Moving Beyond Tropical Infections. PLoS Med. 2008 Feb 26;5(2):e59.
18. Berndt ER, Glennerster R, Kremer MR, Lee J, Levine R, Weizsäcker G, et al. Advance market commitments for vaccines against neglected diseases: estimating costs and effectiveness. Health Econ. 2007;16(5):491–511.
19. Kwabena Tetteh E. Advance Market Commitments for R&D in Diseases That Disproportionately Affect Low-Income Countries. J World Intellect Prop. 2012;15(4):280–303.
20. Leoni PL. Advance Market Commitment: Some Issues and a Bad Remedy [Internet]. Rochester, NY: Social Science Research Network; 2012 Jul [cited 2013 Nov 30]. Report No.: ID 2104110. Available from: http://papers.ssrn.com/abstract=2104110
21. Hughes B. Priority voucher flops. Nat Biotechnol. 2011 Nov;29(11):958–958.
22. Robertson AS, Stefanakis R, Joseph D, Moree M. The Impact of the US Priority Review Voucher on Private-Sector Investment in Global Health Research and Development. PLoS Negl Trop Dis. 2012 Aug 28;6(8):e1750.
23. Pratt B, Loff B. Linking research to global health equity: The contribution of product development partnerships to access to medicines and research capacity building. Am J Public Health. 2013;103(11):1968–78.
24. Morel CM. Reaching Maturity — 25 Years of the TDR. Parasitol Today. 2000 Dec;16(12):522–8.
25. WHO | TDR at a glance [Internet]. WHO. [cited 2012 Mar 1]. Available from: http://www.who.int/tdr/publications/about-tdr/strategy/tdr-glance/en/index.html
26. Stevens AJ, Jensen JJ, Wyller K, Kilgore PC, Chatterjee S, Rohrbaugh ML. The Role of Public-Sector Research in the Discovery of Drugs and Vaccines. N Engl J Med. 2011;364(6):535–41.
27. Frye S, Crosby M, Edwards T, Juliano R. US academic drug discovery. Nat Rev Drug Discov. 2011 Jun;10(6):409–10.
28. Coles LD, Cloyd JC. The Role of Academic Institutions in the Development of Drugs for Rare and Neglected Diseases. Clin Pharmacol Ther. 2012 Aug;92(2):193–202.
29. Moran M, Guzman J, Ropars AL, Illmer A. The role of Product Development Partnerships in research and development for neglected diseases. Int Health. 2010 Jun 1;2(2):114–22.
30. WHO | African trypanosomiasis (sleeping sickness) [Internet]. [cited 2010 Dec 2]. Available from: http://www.who.int/mediacentre/factsheets/fs259/en/
31. La Greca F, Magez S. Vaccination against trypanosomiasis. Hum Vaccin. 2011 Nov 1;7(11):1225–33.
32. Antigenic variation in African trypanosomiasis: a Memorandum. Bull World Health Organ. 1977;55(6):703–13.
33. Nwaka S, Hudson A. Innovative lead discovery strategies for tropical diseases. Nat Rev Drug Discov. 2006 Nov;5(11):941–55.

34. Kneller R. The importance of new companies for drug discovery: origins of a decade of new drugs. Nat Rev Drug Discov. 2010 Nov;9(11):867–82.
35. Powers JB. Commercializing Academic Research: Resource Effects on Performance of University Technology Transfer. J High Educ. 2003;74(1):26–50.
36. Freeman MW, Dervan AP. The Path From Bench to Bedside: Considerations Before Starting the Journey. J Investig Med June 2011. 2011;59(5):746–51.
37. Peng X. University spin-offs: Opportunity or challenge? Nat Mater. 2006 Dec;5(12):923–5.
38. BIO Ventures for Global Health. Developing New Drugs & Vaccines for Neglected Diseases of the Poor: The Product Developer Landscape [Internet]. 2012 Mar. Available from: http://www.bvgh.org/LinkClick.aspx?fileticket=h6a0cJK9drg%3d&tabid=91
39. Adams DJ, Sparrow JC. Enterprise for life scientists: developing innovation and entrepeneurship in the biosciences. Bloxham: Scion; 2008.
40. Endorsements of the London Declaration | Uniting to Combat NTDs [Internet]. [cited 2014 Jan 31]. Available from: http://www.unitingtocombatntds.org/endorsements
41. MRC. Neglected Tropical Diseases (NTDs) highlight notice July 2011 [Internet]. [cited 2014 Jan 30]. Available from: http://www.mrc.ac.uk/Fundingopportunities/Highlightnotices/NTDs/MRC008039
42. MRC. IIB programme [Internet]. [cited 2014 Jan 30]. Available from: http://www.mrc.ac.uk/Ourresearch/Boardpanelsgroups/IIB/Programme/index.htm
43. USAID's Neglected Tropical Disease Program: About the NTD Program [Internet]. [cited 2014 Jan 30]. Available from: http://www.neglecteddiseases.gov/about/index.html#support
44. Improving the health of poor people in developing countries - Policy - GOV.UK [Internet]. [cited 2014 Jan 30]. Available from: https://www.gov.uk/government/policies/improving-the-health-of-poor-people-in-developing-countries
45. Preventing and treating neglected tropical diseases - Improving the health of poor people in developing countries - Policies - GOV.UK [Internet]. [cited 2014 Jan 30]. Available from: https://www.gov.uk/government/policies/improving-the-health-of-poor-people-in-developing-countries/supporting-pages/preventing-and-treating-neglected-tropical-diseases
46. What We Do - Bill & Melinda Gates Foundation [Internet]. [cited 2014 Jan 30]. Available from: http://www.gatesfoundation.org/what-we-do
47. Neglected Infectious Diseases - Bill & Melinda Gates Foundation [Internet]. [cited 2014 Jan 30]. Available from: http://www.gatesfoundation.org/What-We-Do/Global-Health/Neglected-Infectious-Diseases
48. How We Make Grants - Bill & Melinda Gates Foundation [Internet]. [cited 2014 Jan 30]. Available from: http://www.gatesfoundation.org/How-We-Work/General-Information/How-We-Make-Grants
49. The Strategic Plan 2010-20 | Wellcome Trust [Internet]. [cited 2014 Jan 30]. Available from: http://www.wellcome.ac.uk/About-us/Strategy/index.htm
50. Global health trials scheme | Wellcome Trust [Internet]. [cited 2014 Jan 30]. Available from: http://www.wellcome.ac.uk/Funding/Biomedical-science/Funding-schemes/Strategic-awards-and-initiatives/WTX059944.htm
51. Pathfinder Awards | Wellcome Trust [Internet]. [cited 2014 Jan 30]. Available from: http://www.wellcome.ac.uk/Funding/Technology-transfer/Awards/Pathfinder-Awards/index.htm?utm_source=Adestra&utm_medium=email&utm_content=Pathfinder+Awards&utm_campaign=12_10_11+TT+news+issue+12+JF&utm_term=a.c.caplan%40adm.leeds.ac.uk
52. Schervish PG. Major donors, major motives: The people and purposes behind major gifts. New Dir Philanthr Fundrais. 2005;2005(47):59–87.
53. DNDi. Partners [Internet]. [cited 2014 Jan 30]. Available from: http://www.dndi.org/partnership/partners.html#CROs

54. iThemba [Internet]. [cited 2014 Jan 30]. Available from: http://www.ithembapharma.com/index.php
55. Drugs for Neglected Diseases Initiative. Oxaborole SCYX-7158 [Internet]. [cited 2014 Feb 4]. Available from: http://www.dndi.org/diseases-projects/portfolio/oxaborole-scyx-7158.html
56. Von Krogh G, Battistini B, Pachidou F, Baschera P. The changing face of corporate venturing in biotechnology. Nat Biotechnol. 2012 Oct;30(10):911–5.
57. The END Fund Launches to Combat Neglected Tropical Diseases in Africa - The Legatum Group [Internet]. [cited 2014 Jan 30]. Available from: http://www.legatum.com/article/The-END-Fund-Launches-to-Combat-Neglected-Tropical-Diseases-in-Africa
58. Kala-azar [Internet]. OneWorld Health. [cited 2013 Nov 28]. Available from: http://www.oneworldhealth.org/kala-azar
59. Davidson RN, Boer M den, Ritmeijer K. Paromomycin. Trans R Soc Trop Med Hyg. 2009 Jul 1;103(7):653–60.
60. BIO Ventures for Global Health. The Global Health Primer 2012 Snapshot [Internet]. The Global Health Primer 2012 Snapshot. [cited 2014 Jan 31]. Available from: http://www.bvgh.org/LinkClick.aspx?fileticket=PzndmHX3BSU%3d&tabid=91
61. Fexinidazole [Internet]. [cited 2013 Nov 28]. Available from: http://www.dndi.org/diseases-projects/portfolio/fexinidazole.html
62. Torreele E, Bourdin Trunz B, Tweats D, Kaiser M, Brun R, Mazué G, et al. Fexinidazole – A New Oral Nitroimidazole Drug Candidate Entering Clinical Development for the Treatment of Sleeping Sickness. PLoS Negl Trop Dis. 2010 Dec 21;4(12):e923.
63. Kaiser M, Bray MA, Cal M, Trunz BB, Torreele E, Brun R. Antitrypanosomal Activity of Fexinidazole, a New Oral Nitroimidazole Drug Candidate for Treatment of Sleeping Sickness. Antimicrob Agents Chemother. 2011 Dec 1;55(12):5602–8.
64. Drugs for Neglected Diseases Initiative. New drugs for neglected diseases: New hope for forgotten patients [Internet]. [cited 2014 Jan 31]. Available from: http://r4d.dfid.gov.uk/PDF/Outputs/DNDI/DNDi_Europe_brochure.pdf
65. Chesbrough H, Schwartz K. Innovating Business Models with Co-development Partnerships. Res-Technol Manag. 2007;50(1):55–9.
66. Moore GA. Crossing the chasm: marketing and selling technology products to mainstream customers. Oxford: Capstone; 2000.
67. Bray MJ, Lee JN. University revenues from technology transfer: Licensing fees vs. equity positions. J Bus Ventur. 2000 Sep;15(5–6):385–92.
68. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. J Health Econ. 2003 Mar;22(2):151–85.
69. Reinventing Life Science Startups–Therapeutics and Diagnostics | Steve Blank [Internet]. [cited 2014 Feb 6]. Available from: http://steveblank.com/2013/08/19/reinventing-life-science-startups-evidence-based-entrepreneurship/
70. Research C for DE and. Information for Consumers (Drugs) - FDA's Drug Review Process: Continued [Internet]. [cited 2014 Feb 6]. Available from: http://www.fda.gov/Drugs/ResourcesForYou/Consumers/ucm289601.htm
71. DiMasi JA, Feldman L, Seckler A, Wilson A. Trends in Risks Associated With New Drug Development: Success Rates for Investigational Drugs. Clin Pharmacol Ther. 2010 Mar;87(3):272–7.
72. Kalamas J, Pinkus G. The optimum time for drug licensing. Nat Rev Drug Discov. 2003 Sep;2(9):691–2.
73. Elvidge S. The changing shape of dealmaking [Internet]. Available from: http://www.nature.com/biopharmadealmakers/pdf/feature_dealmaking_web.pdf
74. Coller BS, Califf RM. Traversing the Valley of Death: A Guide to Assessing Prospects for Translational Success. Sci Transl Med. 2009 Dec 9;1(10):10cm9.

75. Auerswald PE, Branscomb LM. Valleys of Death and Darwinian Seas: Financing the Invention to Innovation Transition in the United States. J Technol Transf. 2003 Aug 1;28(3-4):227–39.
76. Therapeutics for Rare and Neglected Diseases [Internet]. [cited 2014 Feb 3]. Available from: http://www.ncats.nih.gov/research/rare-diseases/trnd/trnd.html
77. Danzon PM, Nicholson S, Pereira NS. Productivity in pharmaceutical–biotechnology R&D: the role of experience and alliances. J Health Econ. 2005 Mar;24(2):317–39.
78. Respira Design: Complex Requirements for a Simple Solution : Center for Social Innovation (CSI) [Internet]. [cited 2013 Dec 2]. Available from: http://csi.gsb.stanford.edu/respira-design-complex-requirements-simple-solution
79. Tsaioun K, Jacewicz M. De-risking drug discovery with ADDME -- avoiding drug development mistakes early. Altern Lab Anim ATLA. 2009 Sep;37 Suppl 1:47–55.
80. Towse A, Kettler H. Advance price or purchase commitments to create markets for treatments for diseases of poverty: lessons from three policies. Bull World Health Organ. 2005 Apr;83(4):301–7.
81. Miller BLK. Financing the “Valley of Death” : an evaluation of incentive schemes for global health businesses [Internet] [Thesis]. Massachusetts Institute of Technology; 2009 [cited 2013 Nov 30]. Available from: http://dspace.mit.edu/handle/1721.1/54591
82. Richard Gold E, Morin J-F. Promising trends in access to medicines. Glob Policy. 2012;3(2):231–7.
83. Commissioner O of the. Speeding Access to Important New Therapies - Fast Track, Breakthrough Therapy, Accelerated Approval and Priority Review [Internet]. [cited 2014 Feb 3]. Available from: http://www.fda.gov/forconsumers/byaudience/forpatientadvocates/speedingaccesstoimportantnewtherapies/ucm128291.htm
84. Timeline - Pneumococcal AMC - Funding & finance - GAVI Alliance [Internet]. [cited 2014 Feb 3]. Available from: http://www.gavialliance.org/funding/pneumococcal-amc/timeline/
85. Prizes to stimulate innovation | Knowledge Ecology International [Internet]. [cited 2013 Nov 30]. Available from: http://keionline.org/prizes
86. Hubbard T, Love J. A New Trade Framework for Global Healthcare R&D. PLoS Biol. 2004 Feb 17;2(2):e52.
87. Musselwhite LW, Maciag K, Lankowski A, Gretes MC, Wellems TE, Tavera G, et al. First Universities Allied for Essential Medicines (UAEM) Neglected Diseases and Innovation Symposium. Am J Trop Med Hyg. 2012 Jan 1;86(1):65–74.
88. Knowledge Ecology International. KEI Research Note 2008:1 [Internet]. [cited 2014 Feb 3]. Available from: http://keionline.org/misc-docs/research_notes/kei_rn_2008_1.pdf
89. Innovation Management | Crowdsourcing | Challenges | Competitions [Internet]. [cited 2013 Dec 1]. Available from: http://www.innocentive.com/
90. XPRIZE | Making the Impossible Possible [Internet]. [cited 2014 Feb 3]. Available from: http://www.xprize.org/
91. Prize4Life - Prize Model [Internet]. [cited 2013 Dec 1]. Available from: http://www.prize4life.org/page/prizes
92. Big prizes for big challenges [Internet]. [cited 2013 Dec 1]. Available from: http://www.nesta.org.uk/news/13-predictions-2013/big-prizes-big-challenges
93. Wickware P. Resurrecting the resurrection drug. Nat Med. 2002 Sep;8(9):908–9.
94. IDRI: Neglected Disease R&D with a Nonprofit Model : Center for Social Innovation (CSI) [Internet]. [cited 2013 Dec 1]. Available from: http://csi.gsb.stanford.edu/idri-neglected-disease-rd-nonprofit-model
95. Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, et al. Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America. Clin Infect Dis. 2009 Jan 1;48(1):1–12.

96. CDC-Centers for Disease Control and Prevention. Threat Report 2013 | Antimicrobial Resistance | CDC [Internet]. [cited 2013 Nov 29]. Available from: http://www.cdc.gov/drugresistance/threat-report-2013/
97. Report Forecasts Strong MRSA Therapeutics Market [Internet]. 2012 [cited 2014 Feb 4]. Available from: http://www.infectioncontroltoday.com/news/2012/06/report-forecasts-strong-mrsa-therapeutics-market.aspx
98. Research and Markets: Clostridium difficile Infection (CDI) Therapeutics - Pipeline Assessment and Market Forecasts to 2019 | Business Wire [Internet]. 2012 [cited 2014 Feb 4]. Available from: http://www.businesswire.com/news/home/20120510005071/en/Research-Markets-Clostridium-difficile-Infection-CDI-Therapeutics
99. Sharma A, Jacob A, Tandon M, Kumar D. Orphan drug: Development trends and strategies. J Pharm Bioallied Sci. 2010;2(4):290–9.
100. Meekings KN, Williams CSM, Arrowsmith JE. Orphan drug development: an economically viable strategy for biopharma R&D. Drug Discov Today. 2012 Jul;17(13-14):660–4.
101. Fishman MC. Power of Rare Diseases: Found in Translation. Sci Transl Med. 2013 Sep 4;5(201):201ps11–201ps11.
102. Sasinowski FJ. Quantum of Effectiveness Evidence in FDA's Approval of Orphan Drugs. Drug Inf J. 2012 Mar 1;46(2):238–63.
103. Wellman-Labadie O, Zhou Y. The US Orphan Drug Act: Rare disease research stimulator or commercial opportunity? Health Policy. 2010 May;95(2–3):216–28.
104. Villa S, Compagni A, Reich MR. Orphan drug legislation: lessons for neglected tropical diseases. Int J Health Plann Manage. 2009 Mar;24(1):27–42.
105. Commissioner O of the. Designating an Orphan Product: Drug and Biological Products - How to Apply for Designation as an Orphan Product [Internet]. [cited 2014 Feb 5]. Available from: http://www.fda.gov/ForIndustry/DevelopingProductsforRareDiseasesConditions/HowtoapplyforOrphanProductDesignation/ucm135122.htm
106. European Medicines Agency - Human medicines - Orphan designation [Internet]. [cited 2014 Feb 5]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000029.jsp&
107. Shah RR. Regulatory framework for the treatment of orphan diseases. In: Mehta A, Beck M, Sunder-Plassmann G, editors. Fabry Disease: Perspectives from 5 Years of FOS [Internet]. Oxford: Oxford PharmaGenesis; 2006 [cited 2014 Feb 14]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK11567/
108. Perampaladas K, Masum H, Kapoor A, Shah R, Daar AS, Singer PA. The road to commercialization in Africa: lessons from developing the sickle-cell drug Niprisan. BMC Int Health Hum Rights. 2010;10(Suppl 1):S11.
109. Anacor: Neglected Disease R&D Within a For-Profit Model : Center for Social Innovation (CSI) [Internet]. [cited 2013 Dec 2]. Available from: http://csi.gsb.stanford.edu/anacor-neglected-disease-rd-within-profit-model
110. Global Health Innovation Insight Series : Center for Social Innovation (CSI) [Internet]. [cited 2013 Dec 1]. Available from: http://csi.gsb.stanford.edu/global-health-innovation-insight-series
111. Gates Foundation Makes First Equity Investment in a Biotech Startup, Liquidia Technologies [Internet]. Xconomy. [cited 2014 Feb 4]. Available from: http://www.xconomy.com/seattle/2011/03/08/gates-foundation-makes-first-equity-investment-in-a-biotech-startup-liquidia-technologies/
112. Gates Foundation Looks to Make More Equity Bets in Biotech [Internet]. Xconomy. [cited 2013 Dec 3]. Available from: http://www.xconomy.com/seattle/2012/06/28/gates-foundation-looks-to-make-more-equity-bets-in-biotech/

113. Anacor Pharmaceuticals Receives Significant Investment to Support Neglected Disease Research (NASDAQ:ANAC) [Internet]. [cited 2014 Jan 2]. Available from: http://investor.anacor.com/releasedetail.cfm?ReleaseID=754908
114. Robertson MN, Ylioja PM, Williamson AE, Woelfle M, Robins M, Badiola KA, et al. Open source drug discovery - A limited tutorial. Parasitology. 2013 Aug 28;1–10.
115. OSM - Open Source Malaria [Internet]. [cited 2013 Nov 30]. Available from: http://opensourcemalaria.org/#
116. OpenSourceMalaria:Story so far - OpenWetWare [Internet]. [cited 2013 Nov 30]. Available from: http://openwetware.org/wiki/OpenSourceMalaria:Story_so_far
117. Yadav J, Consortium O. Open Source Drug Discovery. Nat Preced [Internet]. 2008 Dec 1 [cited 2013 Nov 30]; Available from: http://precedings.nature.com/documents/2537/version/1%20?iframe=true&width=100%&height=100%
118. Ortí L, Carbajo RJ, Pieper U, Eswar N, Maurer SM, Rai AK, et al. A Kernel for Open Source Drug Discovery in Tropical Diseases. PLoS Negl Trop Dis. 2009 Apr 21;3(4):e418.
119. The Tropical Disease Initiative [Internet]. [cited 2014 Feb 4]. Available from: http://sgt.cnag.cat/TDI/
120. Crowdfunding Early-Stage Drug Development - 10/11/2013 - The Burrill Report [Internet]. [cited 2013 Nov 30]. Available from: http://www.burrillreport.com/article-crowdfunding_early_stage_drug_development.html
121. Orelli B. Biotech crowdfunding paves way for angels. Nat Biotechnol. 2012 Nov;30(11):1020–1020.
122. BIA Citizens' Innovation Funds Second Report - The BioIndustry Association [Internet]. [cited 2014 Feb 3]. Available from: http://www.bioindustry.org/document-library/bia-citizens-innovation-funds-second-report/
123. A Leap Forward to Cure a Neglected Disease [Internet]. [cited 2014 Feb 8]. Available from: https://experiment.com/projects/a-leap-forward-to-cure-a-neglected-disease
124. Experiment - Science Research Crowdfunding [Internet]. [cited 2014 Feb 10]. Available from: https://experiment.com/
125. Catalysts - Funding tool - innovateuk [Internet]. [cited 2014 Feb 10]. Available from: https://www.innovateuk.org/-/catalysts
126. Wiggins J, Gibson DV. Overview of US incubators and the case of the Austin Technology Incubator. Int J Entrep Innov Manag. 2003 Jan 1;3(1):56–66.
127. Becker B, Gassmann O. Corporate Incubators: Industrial R&D and What Universities can Learn from them. J Technol Transf. 2006 Jul 1;31(4):469–83.
128. Waltz E. Start-ups weigh benefits of corporate incubators. Nat Biotechnol. 2008 Mar;26(3):254–5.
129. Neglected Global Diseases Initiative Fund| Connect | Engage | Donate [Internet]. start an evolution UBC | Connect | Engage | Donate. [cited 2013 Dec 1]. Available from: https://startanevolution.ubc.ca/projects/neglected-global-diseases-initiative-fund/
130. 12 digital fundraising trends for 2012 #4 Microdonations [Internet]. Giving in a Digital World. [cited 2014 Feb 10]. Available from: http://givinginadigitalworld.org/2012/01/16/12-digital-fundraising-trends-for-2012-4-microdonations/
131. (RED) [Internet]. The Global Fund to fight AIDS, Tuberculosis and Malaria. [cited 2014 Feb 10]. Available from: http://www.theglobalfund.org/en/partners/privatesector/red/
132. Stop TB Partnership | Massive good can come from new on-line campaign [Internet]. [cited 2013 Dec 1]. Available from: http://www.stoptb.org/news/stories/2010/ns10_008.asp
133. For Ambitious Nonprofits, Capital to Grow [Internet]. Opinionator. [cited 2013 Dec 2]. Available from: http://opinionator.blogs.nytimes.com/2012/06/27/fixes/
134. Raising Funds v. Fundraising: Could social investing change attitudes to profit within the charity sector? - Good Innovation [Internet]. [cited 2013 Dec 2]. Available from:

http://www.goodinnovation.co.uk/news/industry-news/raising-funds-v-fundraising-could-social-investing-change-attitudes-to-profit-within-the-charity-sector-/
135. A guide to charitable bonds. BBC [Internet]. 2011 Dec 21 [cited 2013 Dec 2]; Available from: http://www.bbc.co.uk/news/business-16267298
136. Anderson GF. Spurring New Research For Neglected Diseases. Health Aff (Millwood). 2009 Nov 1;28(6):1750–9.
137. Mackey TK, Liang BA. Global health policy coordination to address neglected tropical diseases. Trop Med Int Health. 2012 Sep;17(9):1053–6.
138. Moon S, Bermudez J, 't Hoen E. Innovation and access to medicines for neglected populations: Could a treaty address a broken pharmaceutical R&D system? PLoS Med. 2012;9(5).